Prediction of Physicochemical Properties For Fe3+ Chelating Agents

Chen, Yu-Lin

Awarding institution:
King's College London

Title: Prediction of Physicochemical Properties For Fe3+ Chelating Agents

Author: Yu-Lin Chen

PREDICTION OF PHYSICOCHEMICAL PROPERTIES FOR Fe^{3+} CHELATING AGENTS

Yu-Lin Chen

Doctor of Philosophy
in
King's College London
University of London

Abstract

Iron is vital for all living creatures but becomes toxic when it exceeds the levels catered for by their natural cellular buffering mechanisms, causing free radical formation *via* the Fenton reaction and ultimately, therefore, leading to oxidative stress. In such situations, for example in repeatedly transfused patients suffering from β-thalassaemia or sickle cell anaemia, iron chelation therapy is required. Desferioxamine (DFO), the most widely used therapeutic chelator, is a hexadentate ligand possessing a very high affinity for Fe^{3+}, but it is not orally active. 3-hydroxypyridin-4-ones (HPOs), the more recently introduced synthetic alternatives, have also been shown to be useful therapeutic chelators, demonstrating the essential affinity and selectivity for Fe^{3+} along with good oral activity. Deferiprone, a typical HPO, has emerged as a prominent therapeutic, able to remove accumulated excess iron from the heart and mitochondria. However, because Deferiprone has some drawbacks of relatively high metabolic instability and a side effect of lowering white blood cell count for a small number of patients, the search continues for other such synthetic chelators with improved properties. In the work reported here, various computational studies have been performed to aid in the rational design of Fe^{3+} chelators, with their physicochemical properties (pK_a, Fe^{3+} affinity, hydration and membrane permeability) predicted by means of quantitative structure property relationship (QSPR) methods, quantum mechanical (QM) calculations, and molecular dynamic (MD) simulations. The pK_a and Fe^{3+} affinity were also studied experimentally with a novel approach devised for dealing with ligands possessing substituents with hydrogen bond donor and/or acceptor groups. The pK_a values predicted using QSPR and QM static calculations (Gaussian 09) were found to differ very significantly from the experimentally determined values. When data from the QM static calculations were combined with a regression model, however, the pK_a predictions were significantly improved, with the predicted values then within ± 0.2 log units of the experimental values, and computing times of the order of 1 day per molecule. These calculations also allowed the determination of possible deprotonation sequences for the predicted compounds. Further pK_a predictions were made by means of QM MD calculations, using Car-Parrinello molecular dynamics (CP-MD). These simulations were found to yield pK_a values within ± 0.3 log units of the experimental values but involved much longer computing times (of the order of 20 days per molecule). In addition, however, the CP-MD simulations also provided valuable insights into the atomistic details of the proton transfer mechanism and the

solvation structure and dynamics at all stages of the reaction. For the three HPOs studied, it was observed that proton transfer takes place along a chain of three H_2O molecules, although direct hydrogen bonds were observed to form transiently. The (Fe^{3+}) log K_1 predictions for the HPOs were made using an entirely novel QM-based methodology (and without knowledge of the chelator pK_a values), yielding log K_1 values within ± 0.32 log units of the experimental values. For the preparation of membrane permeability study, novel Chemistry at HARvard Molecular Mechanics (CHARMM) force fields specifically for use in HPO simulations were developed. These new force fields were validated using Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) MD simulations of the chelators' behaviour in aqueous solution.

Table of Contents

Table of Figures

Table of Schemes

Table of Tables

List of Abbreviations

AD	Alzheimer's disease
BBB	Blood-brain barrier
CHARMM	Chemistry at HARvard Molecular Mechanics
CP-MD	Car-Parrinello molecular dynamics
DFO	Desferioxamine
DFT	Density functional theory
FF	Force field
HF	Hartree-Fock theory
HPO	3-Hydroxypyridin-4-one
HRef	Reference acid
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
MD	Molecular dynamics
MM	Molecular mechanics
PCM	Polarisable continuum model
PD	Parkinson's disease
PES	Potential energy surface
QM	Quantum mechanics
QSAR	Quantitative structure activity relationship
QSPR	Quantitative structure property relationship
RDF	Radial distribution function

$\|M\|$	Mean of absolute deviations from measured values
$\|S\|$	Standard deviation of absolute deviations from measured values
pK_a	Negative logarithmic equilibrium constant of acid
(Fe^{3+}) log K	Logarithmic equilibrium constant between a chelating ligand and Fe^{3+}
(Fe^{3+}) log β	Cumulative logarithmic equilibrium constant between a chelating ligand and Fe^{3+}
pFe^{3+}	$-\log[\text{Free } Fe^{3+}]$ under the conditions that $[Fe^{3+}]_{total}$ is 10^{-6} M and $[\text{Ligand}]_{total}$ is 10^{-5} M at pH 7.4

List of Publications

- Chen, Y. L., Barlow, D. J., Kong, X. L., Ma, Y. M. & Hider, R. C. (2012a). Prediction of 3-hydroxypyridin-4-one (HPO) hydroxyl pKa values. *Dalton Trans.* 41(21): 6549-6557.

- Chen, Y. L., Barlow, D. J., Kong, X. L., Ma, Y. M. & Hider, R. C. (2012b). Prediction of 3-hydroxypyridin-4-one (HPO) log K(1) values for Fe(III). *Dalton Trans.* 41(35): 10784-10791.

- Chen, Y. L., Doltsinis, N. L., Hider, R. C. & Barlow, D. J. (2012c). Prediction of Absolute Hydroxyl pKa Values for 3-Hydroxypyridin-4-ones. *J. Phys. Chem. Lett.*: 2980-2985.

- Ma, Y., Roy, S., Kong, X., Chen, Y., Liu, D. & Hider, R. C. (2012). Design and synthesis of fluorinated iron chelators for metabolic study and brain uptake. *J. Med. Chem.* 55(5): 2185-2195.

- Xu, B., Kong, X. L., Zhou, T., Qiu, D. H., Chen, Y. L., Liu, M. S., Yang, R. H. & Hider, R. C. (2011). Synthesis, iron(III)-binding affinity and in vitro evaluation of 3-hydroxypyridin-4-one hexadentate ligands as potential antimicrobial agents. *Bioorg. Med. Chem. Lett.* 21(21): 6376-6380.

Acknowledgements

I would like to thank Dr David Barlow, my first supervisor, for his guidance, suggestions and patience throughout the whole process of this degree.

I would like to thank Professor Robert Hider, my second supervisor, for providing the knowledge of medicinal chemistry, specifically for iron chelators, and his kindness.

I would like to thank Dr Chris Lorenz for his expertise in classical molecular dynamic simulations.

I would like to thank Dr Nikos Doltsinis for his expertise in quantum mechanical molecular dynamic simulations.

I would like to thank Dr Xiaole Kong for his expertise in stability constants determination.

I would like to thank Dr Yongmin Ma for his expertise in iron chelator synthesis.

I am grateful to my parents, who have fully supported me in every aspect of life, and if it were not for them, I would not have the opportunity to complete this degree.

I hope that this thesis will facilitate future developments of drug discovery and design.

Chapter 1 Introduction

Iron is one of the most abundant elements on Earth and is essential for all living creatures. Due to the nature of its incompletely-filled d-orbitals, a variety of iron valence forms can exist, the most stable of which in aqueous solution are Fe^{2+} and Fe^{3+}. The redox potential between these two forms can be associated with a diversity of metabolic reactions. Although iron is vital for all living cells, it becomes toxic when accumulated in excess. The labile iron can redox cycle between Fe^{2+} and Fe^{3+} *via* the Fenton reaction to generate hydroxyl radicals in the presence of molecular oxygen (Crisponi and Remelli, 2008), thereby giving rise to oxidative stress, which can lead to cell damage (Figure 1-1). In normal individuals, iron levels are controlled tightly and this creates little opportunity to initiate iron-catalysed free radical generation. However, the iron status can vary either systemically as with transfusion-induced iron overload or genetic haemochromatosis, or locally as in ischaemic tissue. In such situations, the elevated iron levels can result in free radical-mediated tissue/organ damage and death (Hoffman, 1991). Thus, iron chelators are required to remove such excess iron. The most widely-used therapeutic iron chelator in haematology over the past 40 years has been Desferrioxamine-B (DFO; Figure 1-2), a hexadentate ligand, which suffers the major drawback of being orally inactive (Hershko *et al.*, 1998). There are numerous different synthetic organic compounds that are able to chelate iron (see Figure 1-3). Ideal iron chelators for clinical use are still searched for and careful design is required to balance their physicochemical properties, such as metal affinity, selectivity and iron-complex stability with their biological attributes of bioavailability and toxicity.

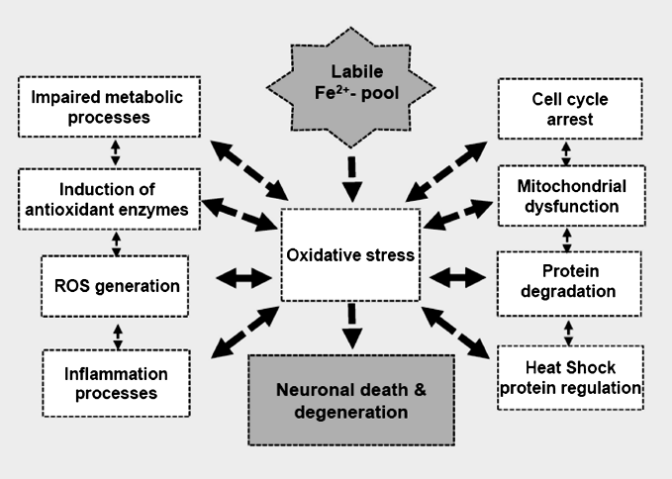


Figure 1-1 Labile iron pool-induced oxidative stress and the many biological functions it may cause to fail (Weinreb *et al.*, 2010).

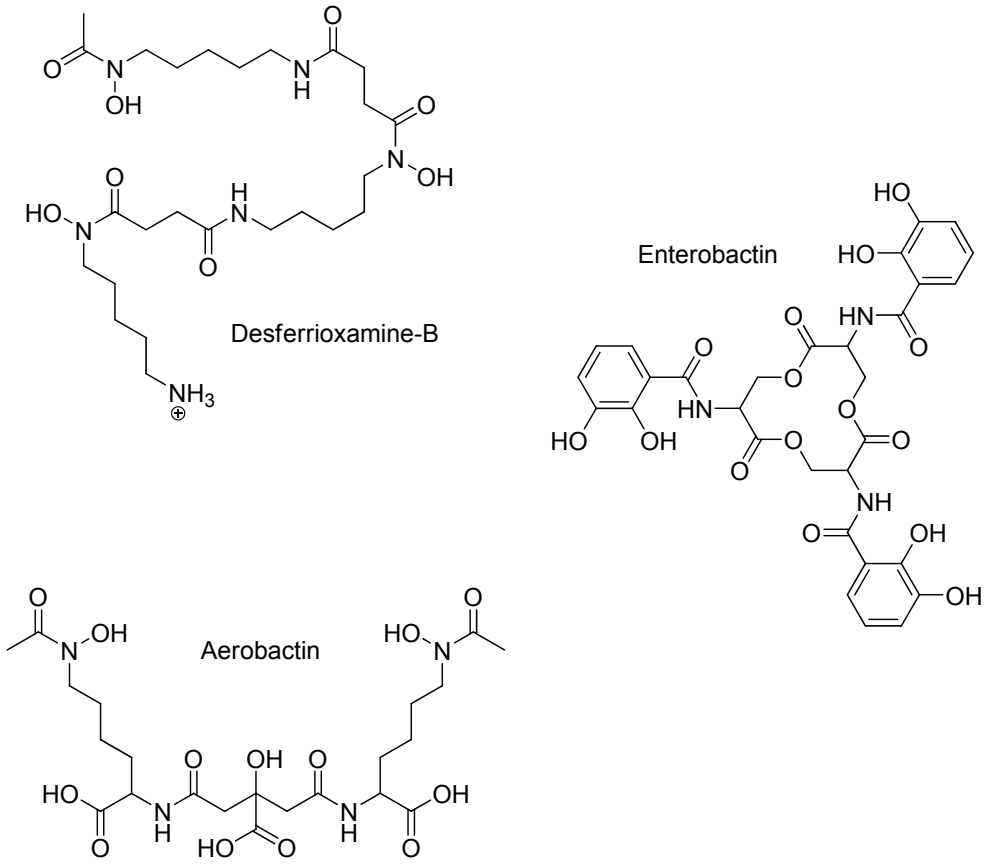


Figure 1-2 Structures of selective siderophores.

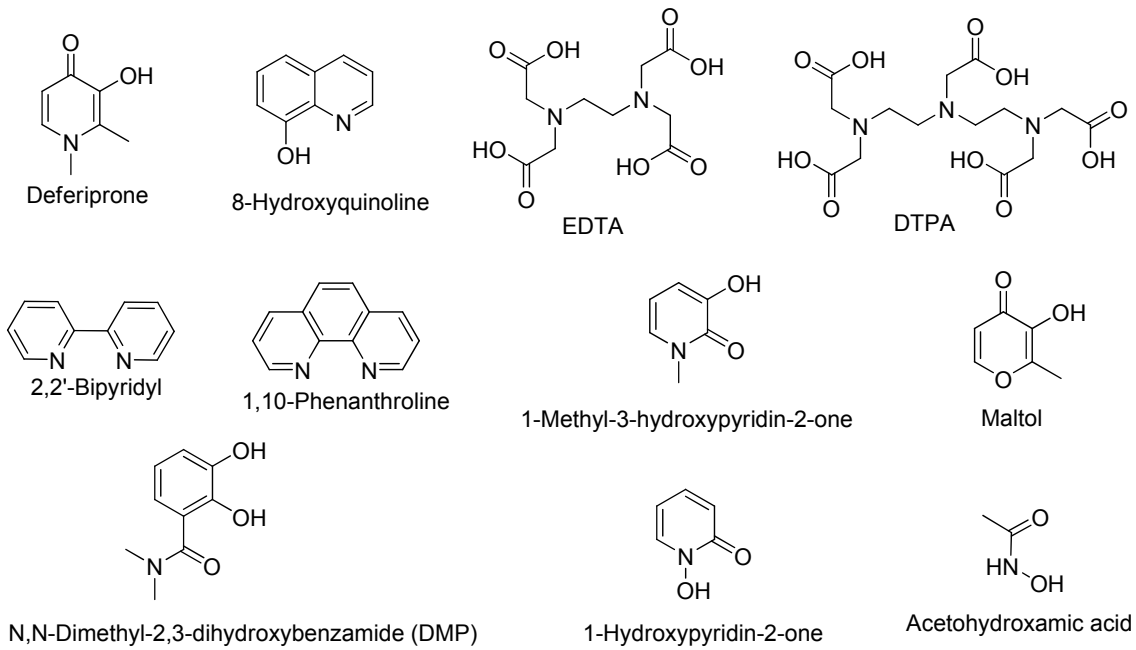


Figure 1-3 Structures of selective chelating agents.

1.1 Iron chelation therapy

1.1.1 Clinical applications

1.1.1.1 Systemic iron overload

Blood transfusion is the most prevalent treatment for inherited haematological diseases such as β-thalassaemia major. Regular blood transfusion can give rise to elevated body iron levels because no physiological excretion routes exist in humans. Each unit of blood (0.40 L) contains approximately 0.25 g of iron and thus, the patient starts to accumulate excess iron when receiving more than two units per year. In the treatment of thalassaemia major, transfusions are normally carried out monthly and thus, more than 2.5 g of iron per patient per year is accumulated in the absence of iron chelation therapy (a normal adult human has a total of 4-5 g of body iron). Excess iron largely accumulates within the liver and the spleen, and if patients are not further treated, death generally takes place in the second decade of life arising from infection or heart disease (Hoffman, 1991). Because iron chelation therapy counteracts the excessive accumulation of iron, thalassaemia major patients can have significantly higher life quality than before. For young patients, bone marrow transplantation can be considered as an alternative treatment with disadvantages of expense and limited availability. As a result, the use of effective orally active iron chelators is much preferred (Porter, 1997).

For the treatment of sickle cell anaemia, transfusion has been introduced more recently because there are advantageous effects to treat crises and to diminish the incidence of stroke (Reed and Vichinsky, 1998). An effective orally active iron chelator, again, allows such patients to receive regular transfusion without the side effect of being iron overloaded. There are a huge number of humans carrying thalassaemia and sickle cell genes in tropical and subtropical regions of the world due to the beneficial effect for heterozygous gene carriers against malaria infection.

Excessive absorption of iron from the daily diet can also cause iron overload. For an adult human, 1 mg of iron per day is normally absorbed under tight control. Several inherited diseases are related to enhanced iron absorption, for instance genetic haemochromatosis which is prevalent in Northern Europe and North America. Iron overload can lead to many pathological symptoms, for example heart failure, liver cirrhosis and diabetes, which commonly appear in the fourth or fifth decade of such patients, when their total body iron level reaches 20-40 g (Barton and Edwards, 2000). Secondary increased iron absorption may also be associated with some

focus of anaemia, particularly in thalassaemia intermedia patients. For such patients, regular blood transfusions may not be necessary to influence their haemoglobin level. Nevertheless, because of increased iron absorption from the gastrointestinal tract, the iron overload-inducing symptoms occur typically in the fourth or fifth decade of their lives (Pippard *et al.*, 1979). Thus, iron chelation therapy can be beneficial for these patients.

Effective iron chelators should possess a pFe^{3+} value (to be discussed in detail in Metal selectivity and affinity section) of approximately 20 (Zhou *et al.*, 2012). If the pFe^{3+} value of an iron chelator is much lower than 20, the chelator is unable to scavenge Fe^{3+} efficiently under physiological conditions. On the contrary, if the pFe^{3+} value is much higher than 20, the chelator is able to compete with iron-binding proteins, such as transferrin, which results in toxic effects.

1.1.1.2 Neurodegeneration (localised iron overload)

In principle, chelation therapy could be useful for the treatment of neurodegenerative diseases related to elevated redox active metal ions in the brain and eye, such as Parkinson's disease (PD), Alzheimer's disease (AD), Friedreich's ataxia, macular degeneration and Hallervorden-Spatz syndrome (Hider *et al.*, 2011). Neurodegeneration is conventionally defined as a progressive loss of function or structure of neurons, including neuron death, in an age-related way. Promising pharmacological targets for the treatment are oxidative stress and protein aggregation (Gaeta and Hider, 2005; Sigel *et al.*, 2006). Oxidative stress is defined as an imbalance between the generation of free radicals and the capability of detoxifying enzymes and antioxidants in the body to reverse them. Oxidative damage occurring in the brain is significant because the brain consumes a high amount of oxygen and contains a relatively low quantity of antioxidants. Furthermore, the brain accumulates redox active metals with age (Crichton and Ward, 2006). Redox active metal ions can be bound to misfolded or partially degraded proteins and thus produce free radicals through redox cycling. This reaction is found to be a critical hallmark in many neurodegenerative diseases (Crichton and Ward, 2006). In order to remove free redox active metal ions in the brain, an adequate chelator must be capable of crossing the blood-brain barrier (BBB) and sequestering the elevated levels of metal to form non-toxic metal complexes, which can be excreted. Again, an effective iron chelator for this purpose should possess a pFe^{3+} value of approximately 20. The ability to permeate the BBB is another crucial factor for the chelator. A molecular weight less than 400 Da and the presence of less than 8 hydrogen bonds are suggested to be ideal properties for this (Pardridge, 2012). In

order to meet this requirement, the choice of chelators is limited and most hexadentate ligands are excluded. However, an extremely large number of bidentate molecules will comply with these limiting guide lines.

1.1.1.3 Bacteriostatic activity

All bacteria and fungi require iron for survival. The bioavailability of iron under aerobic environments is restricted due to the insolubility of $Fe(OH)_3$ and microorganisms have thus evolved the ability to concentrate iron from the environment. One effective means for iron absorption is through the use of siderophores, which are compounds of 500-1500 Da molecular weight, possessing a high affinity and selectivity for Fe^{3+} (Hider and Kong, 2010). There are three main routes of bacteriostatic or antimicrobial activity provided by iron chelating agents:

1. to compete effectively with microbial siderophores for iron and not to be recognised by the cell receptors, thereby competing for the iron supply;

2. to use a siderophore analogue which inhibits the siderophore uptake protein, precluding the microbial siderophores from being transported across the cell membrane, also restricting the iron supply;

3. the “Trojan Horse” approach is to attach an antimicrobial agent to a siderophore backbone, together transported into the cell with its iron-complex. Following the iron release of the complex inside the cell, a series of reactions are triggered to activate the antimicrobial activity of the cargo (Crumbliss and Harrington, 2009).

8-hydroxyquinoline (Figure 1-3) and related compounds were the first chelators to be reported to possess antimicrobial properties more than 50 years ago (Albert *et al.*, 1947; Lowe and Phillips, 1962). N,N’-ethylenebis[2-(2-hydroxyphenyl)-glycine], a hexadentate chelator, has also been found to demonstrate appreciable activity against isolates of pathogenic bacteria and fungi, while EDTA and DTPA (Figure 1-3) exhibit weaker activity (Bergan *et al.*, 2001). EDTA treatment of *P. aeruginosa* biofilms is reported to cause one thousand times more killing than gentamicin treatment (Banin *et al.*, 2006).

1.1.2 Design features of iron chelators with clinical potential

1.1.2.1 Metal selectivity and affinity

In theory, chelators can be designed with preference for either Fe^{3+} or Fe^{2+}. Due to the difference of charge density between Fe^{3+} and Fe^{2+}, the preference of ligating atoms is not the same – Fe^{3+} favours highly charged donors, such as oxygen anions, but Fe^{2+} favours low charged donors, such as nitrogen or sulphur. Ligands in favour of Fe^{2+}, such as 2,2'-bipyridyl and 1,10-phenanthroline (Figure 1-3), can retain appreciable affinities for other biologically important dipositive metal ions, such as Zn^{2+} and Cu^{2+} (Table 1-1). Thus, the design of non-toxic Fe^{2+}-selective chelators is difficult, while Fe^{3+}-selective chelators, typically containing oxygen anions provided by hydroxamates and catecholates, are generally more selective for tripositive metal ions rather than dipositive ones. In addition, most tripositive metal ions, for example Al^{3+} and Ga^{3+}, are not essential for life and thus, Fe^{3+} is regarded as the best target for iron chelator design under physiological conditions. High affinity Fe^{3+} chelators have an additional feature that when chelating Fe^{2+}, the Fe^{2+}-complexes can be rapidly autoxidised to the corresponding Fe^{3+}-complexes under aerobic conditions at physiological pH (Harris and Aisen, 1973).

Table 1-1 Metal affinity constants for selected ligands[a]

Ligand	Log cumulative stability constant					
	Fe^{3+}	Al^{3+}	Ga^{3+}	Cu^{2+}	Zn^{2+}	Fe^{2+}
Desferrioxamine-B	30.6	25.0	27.6	14.1	11.1	7.2
2,2'-Bipyridyl	16.3	–	7.7	16.9	13.2	17.2
1,10-Phenanthroline	14.1	–	9.2	21.4	17.5	21.0
N,N-Dimethyl-2,3-dihydroxybenzamide	40.2	–	–	24.9	13.5	17.5
Acetohydroxamic acid	28.3	21.5	–	7.9	9.6	8.5
Deferiprone	37.2	35.8	32.6	21.7	13.5	12.1
8-Hydroxyquinoline	36.9	33.4	32.9	23	15.8	15.0
EDTA	25.1	16.5	21.0	18.8	16.5	14.3
DTPA	28.0	18.6	25.5	21.6	18.4	16.5

[a] experimental values taken from Martell and Smith (1974-1989). Ligand structures can be found in Figures 1-2 and 1-3.

Because Fe^{3+} is a positively charged ion, most effective iron chelators are negatively charged species, also attracting H^+ and thus, leading to the inevitable competition between Fe^{3+} and H^+. The equilibrium expression ($K_{overall}$) of the formation Fe^{3+}-ligand(L) complex must take into account any pH effect on the reaction.

$$Fe^{3+} + LH_x \rightleftharpoons FeL^{3-x} + xH^+$$

$$K_{overall} = \frac{[FeL^{3-x}][H^+]^x}{[Fe^{3+}][LH_x]}$$

The constant $K_{overall}$ is a pH-dependent conditional equilibrium constant. However, the stability constants found in the literature, for example, the values in Tables 1-1 and 1-2, and the measured values given here in Chapter 2, are pH-independent equilibrium constants (*K*).

$$Fe^{3+} + L^{x-} \rightleftharpoons FeL^{3-x}$$

$$K = \frac{[FeL^{3-x}]}{[Fe^{3+}][L^{x-}]}$$

The constant *K* may be convenient to compare various chelators with different protonation constants (assuming ligands deprotonate completely) in a theoretical sense but this does not account for the important competition between Fe^{3+} and H^+ in aqueous solution, particularly at physiological pH. In order to compare different chelators in a meaningful way, pFe^{3+} is generally adopted as –log[Free Fe^{3+}] under the conditions that $[Fe^{3+}]_{total}$ is 10^{-6} M and $[Ligand]_{total}$ is 10^{-5} M at pH 7.4 (Chaberek and Martell, 1959) (a condition-sensitive indicator, Figure 1-4). This takes into account the impact of ligand denticity, protonation and differences in Fe^{3+}-ligand stoichiometries. For a bidentate ligand, such as deferiprone, the calculation of pFe^{3+} requires at least the five equations (1-5) (the cumulative constant β can be used alternatively, for example $\beta_2 = K_1 \times K_2$; $\beta_3 = K_1 \times K_2 \times K_3$). If the Fe^{3+} affinity of a ligand is not strong, the hydrolysis equilibrium constants of iron need to be included. The lowest pFe^{3+} value for chelators with a weak Fe^{3+} affinity is 14.6, reflecting the Fe^{3+} affinity for OH^- (the mixture of $[Fe(OH)_2]^+$ and $[Fe(OH)_4]^-$ species at pH 7.4, Figure 1-5).

$$Fe^{3+} + L^- \rightleftharpoons FeL^{2+} \qquad K_1 = \frac{[FeL^{2+}]}{[Fe^{3+}][L^-]} \quad (1)$$

$$FeL^{2+} + L^- \rightleftharpoons FeL_2^+ \qquad K_2 = \frac{[FeL_2^+]}{[FeL^{2+}][L^-]} \quad (2)$$

$$FeL_2^+ + L^- \rightleftharpoons FeL_3 \qquad K_3 = \frac{[FeL_3]}{[FeL_2^+][L^-]} \quad (3)$$

$$LH_2^+ \rightleftharpoons LH + H^+ \qquad K_{a1} = \frac{[LH][H^+]}{[LH_2^+]} \quad (4)$$

$$LH \rightleftharpoons L^- + H^+ \qquad K_{a2} = \frac{[L^-][H^+]}{[LH]} \qquad (5)$$

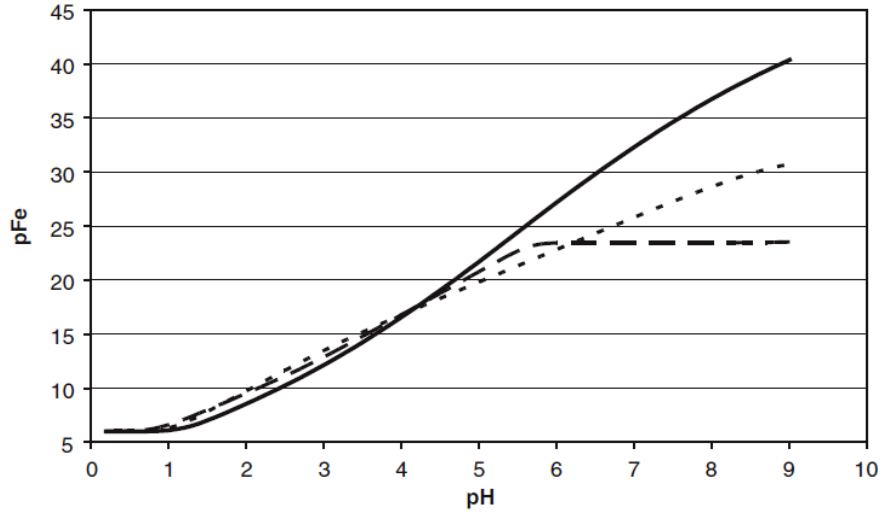


Figure 1-4 pFe^{3+} as a function of pH when $[Fe^{3+}]_{total}$ is 10^{-6} M and $[Ligand]_{total}$ is 10^{-5} M (Valdebenito *et al.*, 2006).
Enterobactin (solid line), aerobactin (dashed line), and desferrioxamine-B (dotted line). The structures can be found in Figure 1-2.

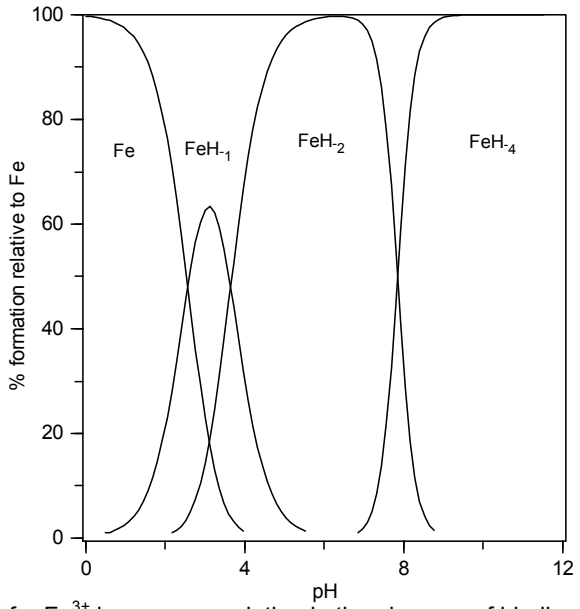


Figure 1-5 Speciation plot for Fe^{3+} in aqueous solution in the absence of binding chelators.

1.1.2.2 Redox activity

As mentioned above, iron can assist to generate free radicals from redox cycling through the Fenton reaction.

$$O_2^{\bullet -} + H_2O_2 \xrightarrow{\text{iron}} O_2 + OH^- + OH^{\bullet} \qquad (6)$$

A therapeutic iron chelator should have an ability to preclude this reaction under physiological conditions and the reduction potential of the iron complex should thus be outside the range of common oxidants and reductants in biological systems (Figure 1-6). The iron complexes of high Fe^{3+}-selective chelators/siderophores possess low reduction potentials, such as enterobactin: -750 mv/NHE, desferrioxamine-B: -468 mv/NHE and deferiprone: -620 mv/NHE (Hider and Kong, 2010; Merkofer *et al.*, 2004). The reduction potentials of iron complexes generally reflect the selectivity and affinity of chelators for Fe^{3+} – the lower the reduction potential, the higher the pFe^{3+} (Figure 1-7). As pFe^{3+} is sensitive to pH, the reduction potential is also pH dependent – the lower the reduction potential, the higher the pH (Crumbliss and Harrington, 2009).

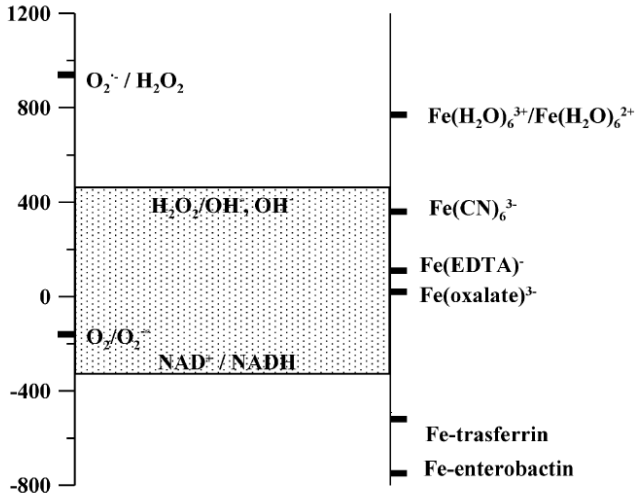


Figure 1-6 Reduction potentials (mV) of important biological reactions and iron-chelator complexes (Crisponi and Remelli, 2008). If the reduction potential of an iron-chelator complex is within the gray area, the complex is susceptible to redox cycling to generate free radicals.

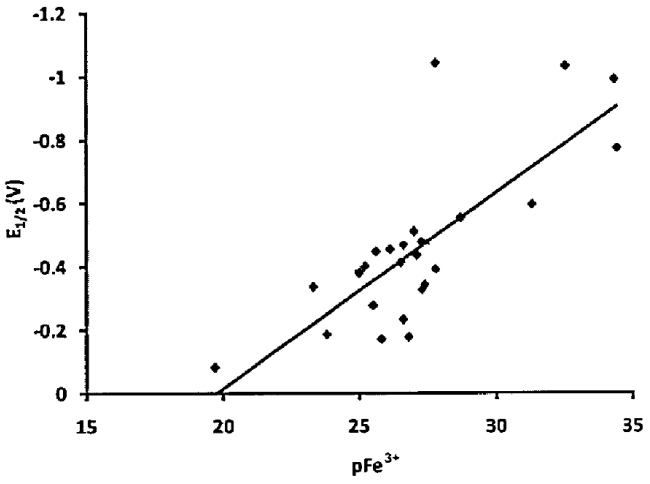


Figure 1-7 Reduction potentials ($E_{1/2}$) as a function of pFe^{3+} values for hexadentate chelators (Crumbliss and Harrington, 2009).

1.1.2.3 Chelator structure

With a view to the prevention of free radical generation, the coordination sites of iron should be completely occupied by chelators to avoid access by oxygen or hydrogen peroxide. The optimal number of coordination atoms for Fe^{3+} is provided by most hexadentate chelators, such as desferrioxamine-B (Figure 1-8) and the iron-complex is relatively more kinetically stable than tridentate and bidentate analogues. This leads to efficient iron sequestration which minimises hydroxyl radical generation. In addition, a lower ligand concentration is required for hexadentate chelators to bind iron. pH-independent equilibrium constants for generic hexadentate, tridentate and bidentate ligands are given in equations (7-9) below:

$$\beta_{\text{hexadentate}} = \frac{[\text{Fe}^{3+}\text{hexadentate}]}{[\text{Fe}^{3+}][\text{hexadentate}]} \qquad (7)$$

$$\beta_{\text{tridentate}} = \frac{[\text{Fe}^{3+}(\text{tridentate})_2]}{[\text{Fe}^{3+}][\text{tridentate}]^2} \qquad (8)$$

$$\beta_{\text{bidentate}} = \frac{[\text{Fe}^{3+}(\text{bidentate})_3]}{[\text{Fe}^{3+}][\text{bidentate}]^3} \qquad (9)$$

Expression of these equilibrium constants is rearranged to the ratio of chelated to unbound iron shown in equations (10-12) below:

$$\beta_{\text{hexadentate}}[\text{hexadentate}] = \frac{[\text{Fe}^{3+}\text{hexadentate}]}{[\text{Fe}^{3+}]} \qquad (10)$$

$$\beta_{\text{tridentate}}[\text{tridentate}]^2 = \frac{[\text{Fe}^{3+}(\text{tridentate})_2]}{[\text{Fe}^{3+}]} \qquad (11)$$

$$\beta_{\text{bidentate}}[\text{bidentate}]^3 = \frac{[\text{Fe}^{3+}(\text{bidentate})_3]}{[\text{Fe}^{3+}]} \qquad (12)$$

In order to obtain the same level of iron scavenging, a higher ligand concentration is required for bidentate and tridentate chelators (assuming equivalent values for $\beta_{\text{hexadentate}}$, $\beta_{\text{tridentate}}$ and $\beta_{\text{bidentate}}$) than for hexadentate chelators. Furthermore, hexadentate chelators generally possess higher pFe^{3+} values than tridentate and bidentate analogues and this can be due to the

preorganisation of hexadentate structures prior to binding iron, minimising the conformational change before and after iron chelation – the lower the conformational freedom of the unbound chelator, the higher the iron-complex stability (Zhou *et al.*, 2012). From a kinetic perspective, chelators can bind or release iron dynamically due to the competition of water, hydroxide ion or other chelating agents in aqueous solution and hexadentate structures constrain the freedom of their chelating groups moving away from bound iron to minimise the dynamics (Crumbliss and Harrington, 2009). For these reasons, the majority of natural siderophores are hexadentate chelators (Hider and Kong, 2010).

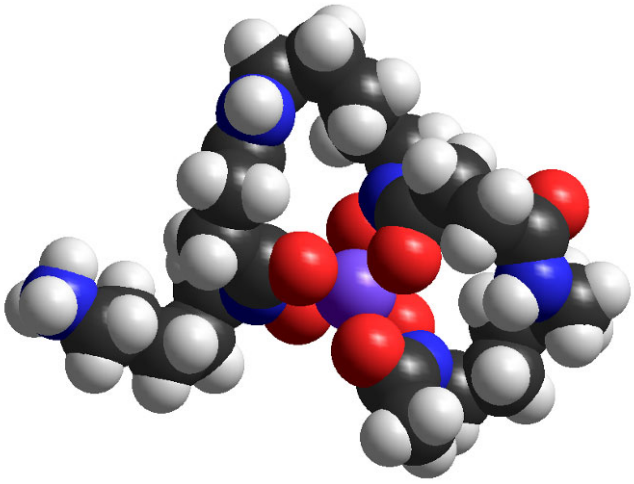


Figure 1-8 Space-filling molecule of iron-desferrioxamine-B crystal structure (a fac, left-hand propeller isomer) (Dhungana *et al.*, 2001).

1.1.2.4 Lipophilicity and molecular weight

With a view to optimising the therapeutic effect for iron chelators, any compound must reach the target sites at sufficient concentration. An orally active iron chelator should thus have an ability to be absorbed from the gastrointestinal tract efficiently and to cross biological membranes, accessing target sites. Three main factors affect drug membrane permeability; these are molecular weight, lipophilicity and ionisation state. Following the Lipinski rule of five (Lipinski *et al.*, 1997), a compound with good absorption or penetration should possess less than 5 hydrogen bond donors, less than 10 hydrogen bond acceptors, a molecular weight less than 500 Da, a calculated logP less than 5. The ionic state of the compound also affects the membrane permeability – uncharged species generally permeate biological membranes faster than charged ones. In order to comply with these conditions, the choice of iron chelators is limited and most hexadentate ligands are excluded due to their molecular weights being more

than 500 Da. Bindentate and tridentate ligands are superior candidates for higher absorption efficiency because of their much lower molecular weights.

1.1.2.5 Enzyme inhibition

Haem-containing enzymes are not generally directly inhibited by iron chelators because their porphyrin groups sequester iron tightly. In contrast, non-haem iron-containing enzymes, for example the lipoxygenase, aromatic hydroxylase and ribonucleotide reductase, can be inhibited by iron chelators (Hider, 1995; Hider *et al.*, 1992). For hydroxypyridinones, more hydrophobic derivatives tend to inhibit lipoxygenases (Abeysinghe *et al.*, 1996) – when the size of the 1-alkyl substitution at the pyridinone ring is increased in a linear manner but it is less evident for the increased size of the 2-substituents (Liu *et al.*, 2002). Indeed, changes in the inhibitory properties of hydroxypyridinones related to various 2-substituents depends more on the size of the substituent than the lipophilicity of the structure (Figure 1-9). The introduction of a hydrophilic 2-substituent for hydroxypyridinones reduces the inhibition significantly, presumably due to a steric effect precluding optimal reaction at the enzyme iron-binding site (Liu *et al.*, 2002). Clearly, iron chelators can be designed to minimise inhibitory effects on a majority of metalloenzymes through physicochemical property modifications.

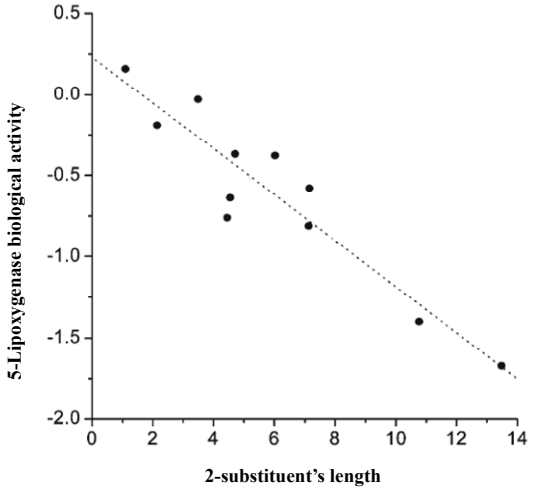


Figure 1-9 Relationship between the hydroxypyridinones with different 2-substituents and the inhibition of 5-lipoxygenase (Liu *et al.*, 2002).

The biological activity = log [P (1-P)$^{-1}$], where P is the enzyme percentage inhibition.

1.1.2.6 Ligand selection for iron(III)

Catechols

Catechol structures possess a high Fe^{3+} affinity due to the high electron density of the two hydroxyl oxygens. This can also be related to the high affinity for H^+ (pK_a values, 12.1 and 8.4) and thus, the cation affinity of catechol is highly pH sensitive (Hider *et al.*, 1981). For bidentate catechols, the 2 : 1 Fe^{3+}-complex is the dominant form at physiological pH and thus, oxidants and reductants can access the iron resulting in redox cycling.

Hydroxamates

Hydroxamate structures possess a lower Fe^{3+} affinity than catechols but the metal ion selectivity is in favour of tripositive over dipositive cations (Table 1-1). The pK_a values of hydroxamates (~ 9) are lower than for catechols and thus H^+ competition at pH 7.4 is less pronounced. As a result, for bidentate hydroxamates, the 3 : 1 Fe^{3+}-complex is the dominant form at physiological pH when the ligand concentration is sufficient. However, the Fe^{3+} affinity of bidentate hydroxamates is insufficient to scavenge iron at pH 7.4 at clinically achievable ligand concentrations and so only tetradentate and hexadentate hydroxamates are effective iron sequesters under physiological conditions.

Hydroxypyridinones

Hydroxypyridinone structures possess balanced features between catechols and hydroxamates and also prefer to coordinate tripositive metal ions with two oxygen atoms. Three classes of hydroxypyridinones capable of acting as bidentate ligand exist, namely 1-hydroxypyridin-2-one, 3-hydroxypyridin-2-one and 3-hydroxypyridin-4-one. Their Fe^{3+} affinities correspond to the pK_a values of their chelating oxygen atoms – the higher the Fe^{3+} affinity, the higher the pK_a value (Table 1-2). 3-Hydroxypyridin-4-ones possess the highest Fe^{3+} affinity in these three classes and are selective for tripositive metal ions (Table 1-1). The carbonyl oxygen possesses significantly higher pK_a value than the other two classes, resulting from the delocalisation of an electron lone pair associated with the aromatic nitrogen (Figure 1-10). This class of compounds are neutral at pH 7.4, and their 3 : 1 Fe^{3+}-complexes are also neutral and the dominant form at physiological pH. This feature assists both the ligand and the Fe^{3+}-complex to cross biological membranes readily. 3-Hydroxypyridin-4-ones possess lower pH-independent log β_3 values than

catechols but the corresponding pFe^{3+} values are higher (Table 1-2). This is mainly because the 3-hydroxypyridin-4-ones possess lower pK_a values.

Table 1-2 pK_a values and Fe^{3+} affinity constants for dioxo-bidentate ligands

Ligand	pK_{a1}	pK_{a2}	Log β_3	pFe^{3+}
N,N-Dimethyl-2,3-dihydroxybenzamide (DMB)	8.4	12.1	40.2	14.7
Acetohydroxamic acid	–	9.4	28.3	14.7
2-Methyl-3-hydroxypyran-4-one (maltol)	–	8.7	28.5	15.3
1-Hydroxypyridin-2-one	–	5.8	27.0	17.5
1-Methyl-3-hydroxypyridin-2-one	0.2	8.6	32.0	18.9
1,2-Dimethyl-3-hydroxypyridin-4-one (deferiprone)	3.6	9.9	37.2	20.3

Data sourced from Liu and Hider (2002) but pFe^{3+} values (–log[Free Fe^{3+}] under the conditions that $[Fe^{3+}]_{total}$ is 10^{-6} M and $[Ligand]_{total}$ is 10^{-5} M at pH 7.4) were re-calculated. The structures can be found in Figure 1-3.

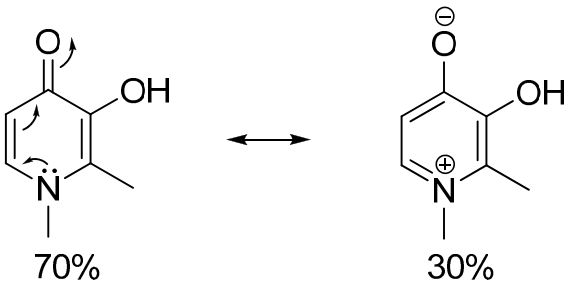


Figure 1-10 Resonance structures of deferiprone (percentages of each species were estimated by Xiao *et al.*, 1992).

1.2 Essential computational chemistry

1.2.1 Structure-property relationships

Structure-property relationships are qualitatively or quantitatively empirically derived relationships between molecular structure and measured properties. In the current literature, this normally implies a quantitative mathematical relationship, frequently obtained using curve fitting programs to optimise a linear combination of molecular information variables for predictions of desired properties. The molecular information variables can be derived from molecular descriptors through simple algorithms or complicated molecular modelling computations. When the desired property is a physical property, the relationship is referred to as a quantitative structure property relationship (QSPR). When the desired property is a biological activity, the relationship is referred to as a quantitative structure activity relationship (QSAR). Due to the use of an empirical relationship, the computing time for predictions is

generally short (typically only requiring a few minutes on a current PC) but the transferability for a compound outside the original training set is questionable.

1.2.2 Force field methods

In force field (FF) methods, the electronic energy is described as a parametric function of the nuclear coordinates and the parameters are fitted to experimental or high level computational data. The basic components of FF methods are atoms, and electrons are not explicitly taken into account. This leads to the requirement of explicit bonding information and the neglect of quantum mechanical nuclear motion. The movement of the atoms is described by classical mechanics, Newton's second law. Molecules are modelled as an ensemble of "balls and springs" with different atom sizes, bond lengths and interaction strengths (Jensen, 2007). These methods are also referred to as molecular mechanics (MM) methods. The basic idea of these methods is the observation that structurally similar molecular units tend to be found in different molecules. For example, CH bond lengths are approximately constant in all molecules (between 1.06 and 1.10 Å). The CH stretch vibrations are also similar (between 2900 and 3300cm^{-1}), relating to the CH bond force constants. The concept that molecules consist of structural units (functional groups) that behave similarly in different molecules is fundamental in organic chemistry. FF methods are in a sense a generalisation of this, with the flexibility of atoms and bonds not being fixed to a single dimension. FF computation additionally predicts the relative energies or barriers for changes in conformation. The calculations of structurally similar molecules are achieved by an appropriate choice of atom types introduced in FF models, which depend on the atomic number and the type of chemical bonding to adjacent atoms.

The FF calculated energy is generally written as

$$E_{FF} = E_{stretch} + E_{bend} + E_{torsion} + E_{vdw} + E_{electro} + E_{cross}$$

where $E_{stretch}$ represents the stretching energy for a bond between two atoms, E_{bend} represents the bending energy for an angle, $E_{torsion}$ represents the torsional energy for a bond rotation, E_{vdw} represents the van der Waals energy, $E_{electro}$ represents the electrostatic energy (E_{vdw} and $E_{electro}$ account for the non-bonded atom–atom interactions) and E_{cross} represents the coupling effect among the first three terms (some MM programs lacking this term) (Figure 1-11). Following this energy function of the nuclear coordinates, optimised relative energies and geometries can be achieved.

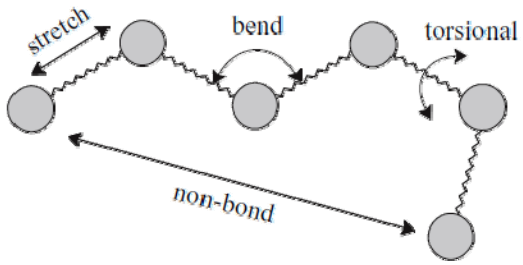


Figure 1-11 Schematic diagram of the "balls and springs" model for force field energy terms (Jensen, 2007).

The simplest function used for $E_{stretch}$ is usually a harmonic oscillator equation, derived from a Taylor expansion.

$$E_{stretch}^{AB} = k^{AB} (R^{AB} - R_o^{AB})^2$$

where A and B represent two different atoms, R represents the distance between two atoms A and B, R_o represents the optimised parameter and k represents the force constant for the AB bond (R_o and k are obtained by fitting to experimental or high level computational data). This form is sufficient for the determination of most equilibrium geometries and simulations at 300K (reasonably accurate up to ~ 8 kcal mol^{-1}) (Jensen, 2007). If higher accuracy is required, the most straightforward modification is to add higher order terms of the Taylor expansion, fitting more parameters.

E_{bend} is the energy for bending an angle and the simplest function is also a harmonic expansion, which is sufficient for most applications.

$$E_{bend}^{ABC} = k^{ABC} (\theta^{ABC} - \theta_o^{ABC})^2$$

where A, B and C represent three different atoms, θ represents the angle between the two bonds AB and BC, θ_o represents the optimised parameter and k represents the force constant for the ABC angle (θ_o and k are obtained by fitting to experimental or high level computational data). By analogy with $E_{stretch}$, higher order Taylor expansion terms can be included for more accurate computation.

$E_{torsion}$ is the energy for rotating a BC bond in a four atoms sequence ABCD, where AB, BC and CD are bonded. The torsional angle is defined as the angle formed between the AB and CD bonds when the system is viewed along the BC bond. The torsional energy is significantly different from $E_{stretch}$ and E_{bend}:

1. The non-bonded terms (E_{vdw} + $E_{electro}$) and the torsional energy can influence rotational barrier together and thus, the torsional parameters are closely related to the non-bonded parameters.

2. Periodicity of the torsional energy function must be complied – the energy should return to the same value when the bond is rotated 360°.

3. The rotation expenditure in energy tends to be low and thus, large deviations from the minimum energy structure may arise. A Taylor expansion in this case is not adequate.

Thus, $E_{torsion}$ is generally written as a Fourier series and the simplest function in an attempt to simulate large systems can be expressed as

$$E_{torsion} = k^{ABCD}(1 + \cos(n\,\Phi^{ABCD} - \delta))$$

where A, B, C and D represent four different atoms, Φ represents the torsional angle, n represents the periodicity of the torsion (for example, n = 1 term means periodicity by 360°, n = 2 term means periodicity by 180°), δ represents a phase factor and k represents the force constant for the ABCD torsional angle (n, δ and k are obtained by fitting to experimental or high level computational data). Higher order terms can be included to modify the detailed shape of the rotation barrier profile. An out-of-plane motion can be described by an improper torsion (Figure 1-12), which is also composed of four atoms and can be controlled by the same function (or a harmonic function).

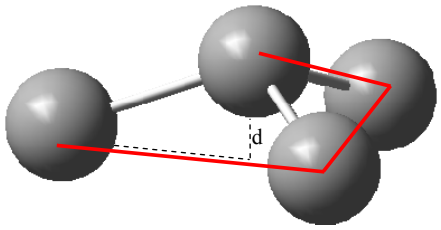


Figure 1-12 Out-of-plane motion described by an improper torsion (red line).

E_{vdw} is the van der Waals energy which describes the attraction or repulsion among non-bonded atoms, as one part of non-bonded energy. This may be explained as the non-polar part of the non-bonded interaction, which is not associated with electrostatic energy arising from atomic charges. This mainly results from induced dipole-dipole (and/or other multipole-multipole) interactions due to a slightly uneven distribution of transient electron movement (electron correlation). E_{vdw} is very positive at short distances, becomes slightly negative to a minimum at

the interatomic contacting distance, and approaches zero as the distance increases. A widely-used function for this is the Lennard-Jones (LJ) potential (Jones, 1924).

$$E_{LJ}(R) = \varepsilon\,[(R_o\,R^{-1})^{12} - 2(R_o\,R^{-1})^{6}]$$

where R_o is the minimum energy distance and ε is the depth of the minimum. The choice of the exponent in the repulsive part to be twelve is only for a computational convenience without theoretical argument. Different programs may have slightly modified forms derived from the above function.

$E_{electro}$ is the other part of non-bonded energy due to internal distribution of the electrons, bringing about positive and negative molecular parts. For example, a carbonyl group possesses a positively charged carbon and a negatively charged oxygen. This can be modelled through the assignment of partial charges to each atom or the assignment of dipole moments to each bond. These two means can produce similar results and the atomic charge model is preferred by most force fields. The interaction between two point charges is obtained using the Coulomb potential.

$$E_{electro}(R^{AB}) = Q^{A}\,Q^{B}\,(\varepsilon\,R^{AB})^{-1}$$

where Q is a point charge, R is the distance between points A and B and ε is a dielectric constant. The atomic charges are usually assigned through fitting to the electronic potential generated by electronic structure methods.

Most force fields calculate E_{vdw} and $E_{electro}$ for bonded atom pairs separated by three bonds or more and thus, the rotation barrier for the AD pair in an ABCD sequence is determined by $E_{torsion}$, E_{vdw} and $E_{electro}$ terms together. $E_{torsion}$ may be interpreted as a necessary correction for obtaining the correct rotation barrier once the non-bonded interactions have been determined. Some force fields also include E_{vdw} for bonded atoms separated by one atom, called a Urey–Bradley term. In this case, the bending energy for a three atom sequence is a mixture of E_{bend} and E_{vdw}.

Validation of a FF is often made through checking the accuracy of reproduction for reference data, which may or may not be adopted in the actual parameterisation. The direct comparison of different force fields in accuracy is challenging due to their diverse use of reference data. Indeed, no FFs are perfect for universal applications, each of which has its

benefits and drawbacks. They can produce reliable simulated results for compounds that are structurally similar to those in the parameterisation sets, but may not be so for other systems.

1.2.3 Electronic structure methods

The interest of electron distribution in detail must be solved using quantum mechanics (QM). Electrons are very light particles and thus their behaviour cannot be predicted correctly using classical mechanics, even qualitatively. QM is centred on the solution to Schrödinger equation and the time-independent form is written as

$$\mathbf{H}\Psi = E\Psi$$

where **H** represents the Hamiltonian operator, Ψ represents the wave function and E represents an energy function. When the solutions are produced without the use of experimental data, the methods are generally referred to as *ab initio* calculations, in contrast to semi-empirical methods (which use experimental parameters). The Born–Oppenheimer approximation – by which there is the neglect of the association between the nuclear and electronic movement – is the key to solving the Schrödinger equation. This permits the electronic part to be solved as function of the nuclear positions. Thus, the solution to the electronic Schrödinger equation, given a set of nuclear coordinates, requires the main part of computational effort.

The dynamics of a many-electron system is extremely complicated and thus, this requires sophisticated computational methods. Independent particle models – one electron motion is independent of the other electrons – are introduced to simplify the system. This can be achieved either through the neglect of all other electrons but the most significant one, or through the consideration of the average effect accounting for all other electrons. The acceptable accuracy in electronic structure theory can only be obtained by the latter, namely Hartree-Fock (HF) theory. In this theory, each electron is described by an orbital and the total wave function is represented as a product of orbitals. The overall wave function must be antisymmetric (sign change for any two electrons exchange) and this can be achieved through the arrangement of the orbitals in a Slater determinant. The variational principle is used to obtain the optimal set of HF orbitals with the lowest energy under the condition that the wave function is a single Slater determinant. The molecular orbital shape describes the probability of an electron appearance, which includes the effects of the average all other electron repulsions and all nuclei attraction. Because the other electrons are described by their individual orbitals,

the solution to the HF equations must be determined by iteration. When a proper basis set is chosen for the molecular orbitals, the equations can be treated as a matrix for their eigenvalues and eigenvectors. Integrals of one and two-electron operators over basis functions, which are multiplied by density matrix elements, are the elements in the Fock matrix. Thus, the repeated diagonalisations of the Fock matrix can be used to solve the HF equations in a basis set. Based on the HF model, if additional approximations are adopted, this brings out semi-empirical methods. Or if additional determinants are incorporated, this produces models which can be resolved towards the exact solution of the electronic Schrödinger equation (Jensen, 2007).

Semi-empirical methods are derived from the HF model through the neglect of all integrals associated with greater than two nuclei in the original Fock matrix. Such approximations alone will result in a poor model due to the inherently limited accuracy of the HF model. Semi-empirical methods further parameterise the remaining integrals with a reference to experimental data, particularly for molecular geometries and energies. These methods are computationally more cost-effective than the *ab initio* HF method but are limited to the parameterised systems.

HF theory neglects the correlation between electrons and further improvement in accuracy is required to include the effect. Because HF is the optimal single-determinant wave function, methods incorporating the electron correlation require a multi-determinant wave function. Such methods are computationally more expensive but can provide systematically improving results towards the exact solution of the Schrödinger equation.

Density functional theory (DFT) using the Kohn–Sham model (Kohn and Sham, 1965) can be interpreted as an improved method (over HF theory) – the electron correlation effect is modelled as a function of the electron density. DFT is also an independent-particle model but provides significantly superior results with the same computational cost as HF methods. The main drawback of DFT is no systematic approach for improving results towards the exact solution.

1.2.4 Molecular dynamics

Molecular Dynamics (MD) methods produce a series of time-dependent points in a trajectory using Newton's second equation in a series of finite time steps, evolving an initial set of atomic coordinates and velocities. A typical time step is one femtosecond (10^{-15} s) and if a total

simulation time is one nanosecond (10^{-9} s), this already requires a million computational steps. Such simulation time is significantly shorter than many naturally occurring phenomena and MD methods often only sample a part of phase space near the starting point. Moreover, MD methods simulate the physical configuration changes and thus are susceptible to becoming trapped in conformational states associated with local energy minima.

MD simulations must run in Cartesian coordinates, and constraints are somewhat difficult to be enforced in the system. The simulations require small time steps and often compute relatively unimportant bond stretching and angle bending motions with a significant computational expenditure. Also, energy barriers are hard to overcome by MD simulations due to an inherent force being generated to take the system back towards minima and working against movements to states of higher potential energy. In theory, MD is deterministic – two simulations with the same starting conditions should produce the same trajectory. However, because different computers (and compilers) generate different round-off errors, the numerical errors arising in each time step will gradually accumulate to become significant. Thus, in practice, MD simulations are non-deterministic and show chaotic behaviour for timescales greater than fifty picoseconds (10^{-12} s). MD simulations include both atomic velocities and time dependence and thus models for time-dependent events, such as transport and diffusion phenomena, are appropriate. The implementation of MD simulations requires calculations for the energy and the force (first derivative of the energy) on all particles in the system. Parameterised energy functions used in force field methods can compute these two terms readily. Because the motions of all particles are different in each time step, the energy function (and gradient) must be updated accordingly.

The non-deterministic behaviour of MD simulations in practice may be regarded as a potential problem. In fact, the chaotic behaviour may assist in a more complete sampling to obtain a representative sample of the phase space. However, the random and chaotic features in simulations can lead to troublesome identification of errors in programming. The validation of a novel simulation program cannot be made readily through the comparison of exact numbers with another program – the same program run on different types of machines may not generate the same results. The inherent statistical errors in all simulations may cover the part of small systematic errors generated by program errors. Thus, development of simulation packages

must be made carefully through monitoring many different quantities to ensure the validation (van Gunsteren and Mark, 1998).

A simulation can be described by quantities, such as number of particles (N), pressure (P), volume (V), temperature (T), total energy (E), chemical potential (CP), and some of these are correlated with each other. For a system with a constant N, either P or V can be fixed, but not both. Similarly, either the T or E can be fixed, but not both, and a constant CP cannot co-exist with a constant N. The ensemble is labelled following the fixed quantities. A MD simulation conserves energy and thus, is the natural NVE type, but other types of ensembles can also be generated through suitable algorithms.

The results of simulations and experiments require the average over the number of molecules for a period of time but their averaging scales tend to be significantly different from each other. Thus, the direct comparison between the computed quantities and the experiments is not entirely clear. For instance, an IR spectrum records the average over a sample composed of perhaps a thousand molecules for ten femtoseconds. On the other hand, a simulation might simulate the molecular motions of perhaps a hundred molecules for one nanosecond. The assumption of the ergodic hypothesis states that the average over a large number of particles for a short time is equivalent to the average over a small number of particles for a long time – ensemble-averaging is equivalent to time-averaging. In other words, the ergodic hypothesis implies that wherever a system begins, any other point in phase space is possibly reached. MD is a time-averaging method.

1.2.5 Solvation models

Evaluation of the effect on the environment, such as a solvent, is one of important aspects in computational chemistry. Methods for the evaluation of the solvent effect can generally be categorised into two types – 1. explicit solvent models including the individual solvent molecules; 2. implicit solvent models treating the solvent as a continuous medium (Jensen, 2007). Combinations of these two types are also feasible – for instance, first solvation shell consists of explicit solvent molecules and the rest is treated using a continuum model. They may be subdivided based on the use of classical or quantum mechanical computation. The most crucial solvent is water and most developed methods are centred on this. The impacts of solvation can be divided into two major parts: 1. Non-specific (long-range) solvation including polarisation and

dipole orientation; 2. Specific (short-range) solvation including hydrogen bonds, van der Waals interaction, solvent shell structure, solvent–solute dynamics, charge transfer effects and hydrophobic effects (entropy effects). These effects bring out a screening of charge interactions – the (macroscopic) dielectric constant is greater than one. The microscopic interactions, depending on the nature of solvent molecules, mainly occur in the first solvation shell (even though the second solvation shell might also be important for highly charged ions).

In principle, a molecular model using simulation methods with periodic boundary conditions enables one to take into account all of the above solvent effects and the quality of the results depends on the realistic level of the description for the solvent-solute and solvent-solvent interactions. A realistic model requires many explicit solvent molecules and thus, force field methods are generally preferred in practice. Because the use of polarisable force fields is not yet common, there is a significant deficiency in the modelling of non-specific solvation. Quantum mechanical methods, such as Car-Parrinello MD (http://www.cpmd.org/) using density functional theory for the description of the interaction, are substantially more computationally expensive and thus, only a limited sampling of the phase space can be provided. These methods can take into account the polarisation but often provide poor description of van der Waals interactions. Semi-empirical electronic structure methods are generally insufficiently accurate for the calculation of intermolecular potentials.

Non-specific solvation can alternatively be modelled by treating the solvent as a homogeneous medium with a dielectric constant (ε), in which a solute is placed in an adequately shaped cavity, namely continuum solvation models (Roux and Simonson, 1999). The solvation energy can be divided into three parts: 1. cavity generation in the medium requires energy; 2. dispersion/repulsion interactions between the solute and solvent; 3. the induced electrostatic charge interactions between the solute and solvent. There are five main factors affecting the computational results of the models: 1. the size and shape of the cavity; 2. the algorithm of dispersion/repulsion contributions; 3. the algorithm of the solute charge distribution; 4. the molecular or quantum mechanical algorithm for the solute; 5. the algorithm of the dielectric medium (Jensen, 2007). In general, this type of model can mimic non-specific solvation reasonably but cannot mimic the specific solvation which appears in the first solvation sphere.

1.3 Aim of project

To explore the use of quantum or molecular mechanical and molecular dynamic simulations in modelling the pK_a, iron binding affinity and hydration properties of 3-hydroxypyridin-4-ones, with the overall aim of facilitating a rational design of therapeutic Fe^{3+} chelating agents.

Chapter 2 Experimental Fe^{3+} Stability Constants Determination

2.1 Introduction

In the design and clinical development of therapeutic iron chelators, it is crucial to know, or to be able to reliably predict stability constants, which quantify their binding affinity for H^+ and Fe^{3+}. With the aim of obtaining data for prediction validation, the stability constants of thirty-three HPOs and one 8-hydroxyquinoline-2-carboxylic acid (8HQC, a tridentate ligand) were experimentally determined.

2.2 Methods

Stability constants were experimentally determined by use of an automated titration system: 1. sample preparation; 2. calibration of pH meter; 3. load of sample into the system (25 ^{o}C, 0.1 M KCl ionic strength); 4. export of UV spectra data into pHab (Gans *et al.*, 1999) to obtain experimental values; 5. export of experimental values into HySS (Alderighi *et al.*, 1999) to obtain speciation plots and pFe^{3+} values. Full details are provided in Kong (2008) and Kong *et al.*, (2006). The key factors of starting titration pH value and the molar ratios of ligand (L) : metal ion (Fe^{3+}) when determining Fe^{3+} affinity constants were here taken as the HPO carbonyl pK_a value, the lower pK_a value of the two HPO oxygens coordinating Fe^{3+}, minus 2 if possible (according to Henderson-Hasselbalch equation), and a 10 : 1 molar ratio of L : Fe^{3+}. (The reasons for these choices are discussed in Chapter 5.) For HPOs possessing more than 2 pK_a values, the assignment of pK_a values (to the two HPO oxygens, directly related to experimental stability constants for Fe^{3+}) is an additional issue and the QM calculations reported in Chapter 3 proved valuable in this assignment.

2.3 Results

The UV/Vis spectra and corresponding speciation plots for CP20 (a typical HPO with two pK_a values) are presented in Figure 2-1. For pK_a titration, an isosbestic point at λ 290 nm occurs when [LH] of CP20 deprotonates. $[LH_2]^+$ dominates in solution from pH 1 to pH 3. [LH] dominates in solution from pH 5 to pH 8. $[L]^-$ dominates in solution from pH 11 to pH 13. The two pK_a values (9.83 and 3.61) were determined from curve fitting analysis of the spectra. For Fe^{3+} affinity titration, with a 10 : 1 molar ratio of L : Fe^{3+}, two isosbestic points occur in the spectral series: one at λ 613 nm, corresponding to the formation of $[Fe^{3+}L_2]^+$ and the other at λ 506 nm,

corresponding to the formation of $[Fe^{3+}L_3]$. $[Fe^{3+}L_1]^{2+}$ dominates in solution from pH 0.5 to pH 1.5. $[Fe^{3+}L_2]^{+}$ dominates in solution from pH 2 to pH 3. $[Fe^{3+}L_3]$ dominates in solution from pH 4 through to pH 8. The log K_1, log β_2, log β_3 values (15.01, 27.39 and 37.35, respectively) were determined from curve fitting analysis of the spectra. pFe^{3+} (20.59) was determined through export of the stability constants reported above into HySS.

The UV/Vis spectra and corresponding speciation plots for CP511 (a HPO with two pK_a values and strong intramolecular hydrogen bonding (Piyamongkol *et al.*, 2010)) are presented in Figure 2-2. For pK_a titration, two isosbestic points occur in the spectral series: one at λ 299 nm, corresponding to the formation of [LH] and the other one at λ 310 nm, corresponding to the formation of $[L]^{-}$. $[LH_2]^{+}$ dominates in solution from pH 1 to pH 2. [LH] dominates in solution from pH 3 to pH 6. $[L]^{-}$ dominates in solution from pH 7 to pH 9. The two pK_a values (6.71 and 2.42) were determined from curve fitting analysis of the spectra. For Fe^{3+} affinity titration, with a 10 : 1 molar ratio of L : Fe^{3+}, two isosbestic points occur in the spectral series: one at λ 600 nm, corresponding to the formation of $[Fe^{3+}L_2]^{+}$ and the other at λ 505 nm, corresponding to the formation of $[Fe^{3+}L_3]$. $[Fe^{3+}L_1]^{2+}$ dominates in solution from pH 0 to pH 0.5. $[Fe^{3+}L_2]^{+}$ dominates in solution from pH 1 to pH 1.5. $[Fe^{3+}L_3]$ dominates in solution from pH 2 through to pH 8. The log K_1, log β_2, log β_3 values (12.03, 22.38 and 31.26, respectively) were determined from curve fitting analysis of the spectra. pFe^{3+} (21.55) was determined through export of the stability constants reported above into HySS.

The UV/Vis spectra and corresponding speciation plots for CP417 (a HPO with three pK_a values) are presented in Figure 2-3. For pK_a titration, two isosbestic points occur in the spectral series: one at λ 302 nm, corresponding to the formation of [LH] and the other one at λ 305 nm, corresponding to the formation of $[L]^{-}$. $[LH_3]^{2+}$ dominates in solution from pH 0 to pH 1. $[LH_2]^{+}$ dominates in solution from pH 2.5 to pH 5. [LH] dominates in solution from pH 7.5 to pH 10. $[L]^{-}$ dominates in solution from pH 11 to pH 13. The three pK_a values (10.47, 6.71 and 2.03) were determined from curve fitting analysis of the spectra. For Fe^{3+} affinity titration, with a 10 : 1 molar ratio of L : Fe^{3+}, two isosbestic points occur in the spectral series: one at λ 601 nm, corresponding to the formation of $[Fe^{3+}L_2]^{+}$ and the other at λ 503 nm, corresponding to the formation of $[Fe^{3+}L_3]$. $[Fe^{3+}L_1]^{2+}$ dominates in solution from pH 0.3 to pH 0.9. $[Fe^{3+}L_2]^{+}$ dominates in solution from pH 1 to pH 1.8. $[Fe^{3+}L_3]$ dominates in solution from pH 2 through to pH 8. The log K_1, log β_2, log β_3 values (10.98, 20.54 and 28.78, respectively) were determined from curve

fitting analysis of the spectra. pFe^{3+} (19.07) was determined through export of the stability constants reported above into HySS. Members of entire HPO series have similar UV/Vis spectra for their $[Fe^{3+}L_1]^{2+}$, $[Fe^{3+}L_2]^{+}$ and $[Fe^{3+}L_3]$ species. The stability constants of all thirty-three HPOs experimentally determined in this study are presented in Table 2-1 and the additional UV/Vis spectra and corresponding speciation plots are presented in Appendix A.

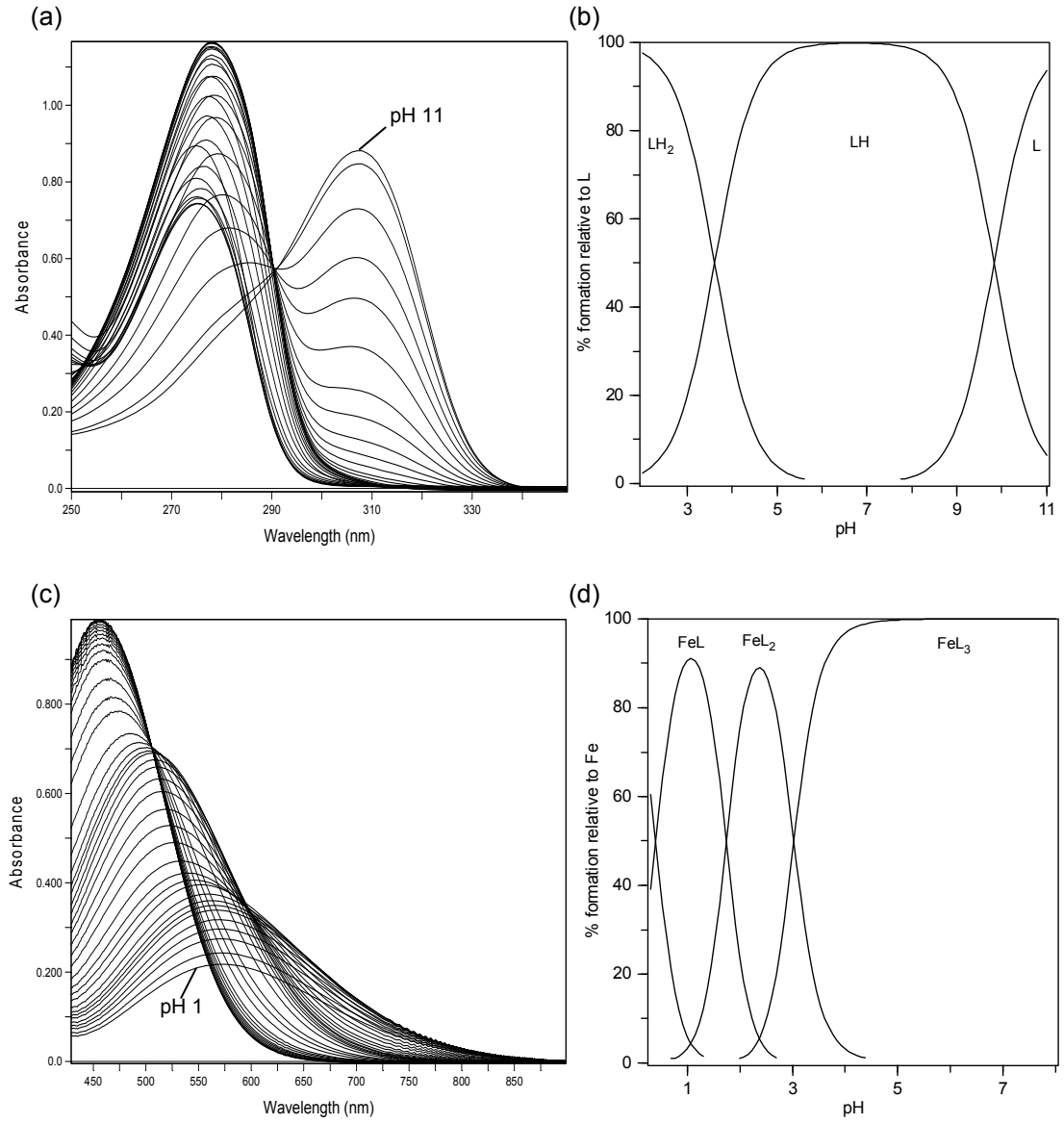


Figure 2-1 UV/Vis spectra and corresponding speciation plots for CP20.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1176.8 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 456.6 μM and [Fe^{3+}] = 46.1 μM over the pH range 1-7.

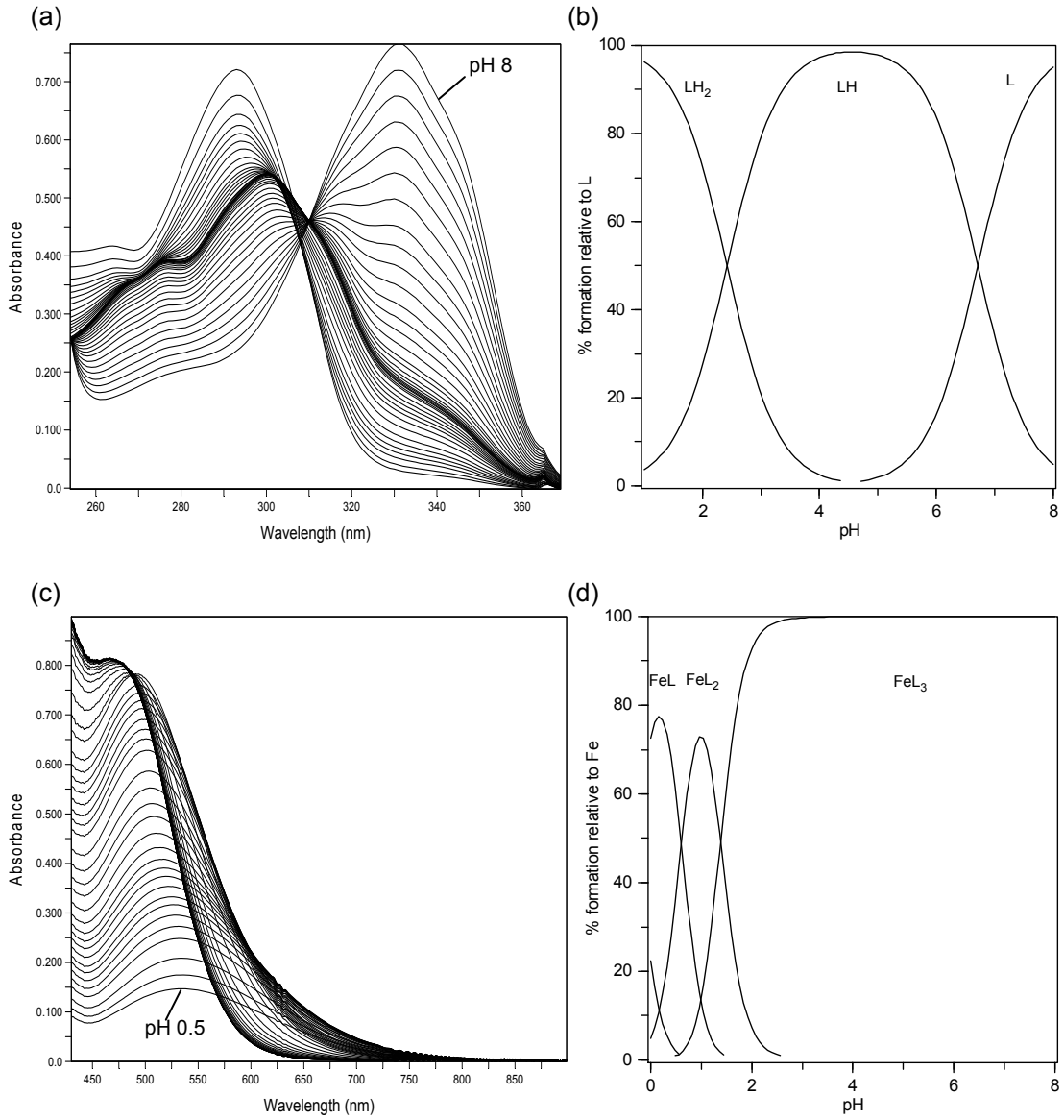


Figure 2-2 UV/Vis spectra and corresponding speciation plots for CP511.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 879.7 μM over the pH range 1-8; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 445.4 μM and [Fe^{3+}] = 44.2 μM over the pH range 0.5-8.

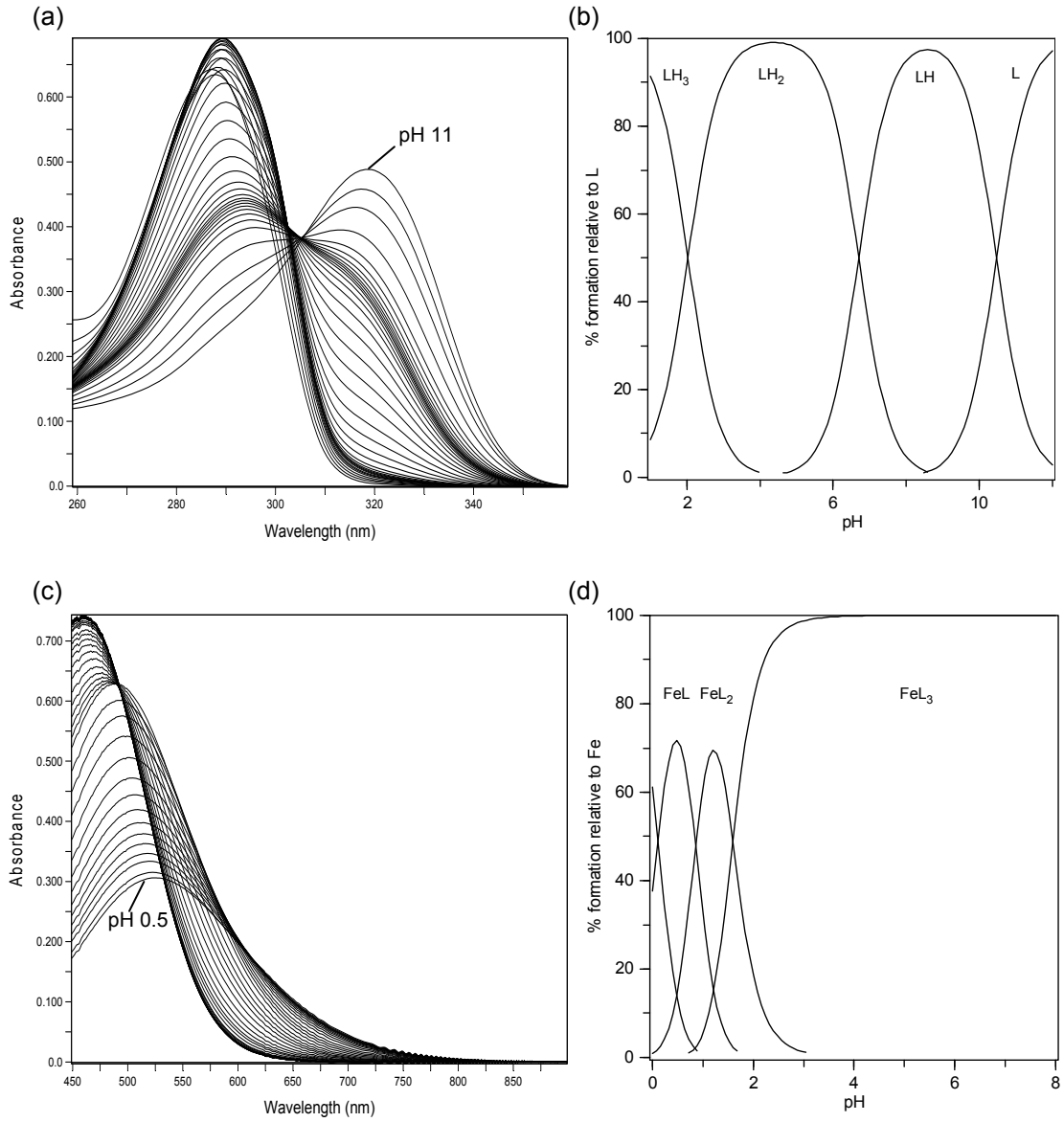
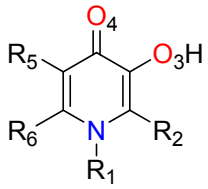


Figure 2-3 UV/Vis spectra and corresponding speciation plots for CP417.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 930.8 μM over the pH range 1-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 363.4 μM and [Fe^{3+}] = 35.5 μM over the pH range 0.5-8.

Table 2-1 Thirty-three HPOs' structures and experimental stability constants.



HPOs	R_1	R_2	R_5	R_6
CP20	$-CH_3$	$-CH_3$	−H	−H
CP28	−H	$-CH_3$	−H	−H
CP38	$-C_2H_4COOH$	$-CH_3$	−H	−H
CP40	$-C_2H_4OH$	$-CH_3$	−H	−H
CP70	−H	−H	−H	$-CH_3$
CP102	$-C_2H_4OH$	$-C_2H_5$	−H	−H
CP110	$-C_2H_4COOH$	$-C_2H_5$	−H	−H
CP111	$-C_3H_6COOH$	$-C_2H_5$	−H	−H
CP352	$-C_2H_4OH$	$-CH(C_6H_5)OH$	−H	$-CH_3$
CP359	$-C_2H_5$	$-CH_2OH$	−H	−H
CP364	$-CH_3$	$-CH_2OH$	−H	−H
CP365	$-C_2H_5$	$-CHCH_3OH$	−H	−H
CP366	$-C_3H_6OH$	$-CH_2OH$	−H	−H
CP370	$-C_2H_5$	$-CH_2OCH_3$	−H	$-CH_3$
CP372	$-C_3H_6OH$	$-CHCH_3OH$	−H	−H
CP374	$-CH_3$	$-CHC_2H_5OH$	−H	$-CH_3$
CP375	$-CH_3$	$-CHC_2H_5OCH_3$	−H	$-CH_3$
CP414	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_3$
CP417	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_2OH$
CP510	$-CH_3$	−H	−H	$-CONHCH_3$
CP511	−H	$-CONHCH_3$	−H	$-CH_3$
CP529	−H	$-CONHCH_3$	−H	−H
CP545	−H	$-CONHC_2H_4OH$	−H	−H
CP616	−H	$-CH_3$	$-CH_3$	−H
CP751	$-CH_3$	$-CH_3$	$-CH_3$	−H
YMF1	−H	$-CHCF_3OH$	−H	−H
YMF3	−H	−F	−F	−F
YMF4	−H	−F	−H	−F
YMF6	−H	−F	−H	−H
YMF13	−H	−F	$-CH_3$	−H
YMMO1	$-CH_3$	$-CH_3$	$-OCH_3$	−H
CN116	$-CH[CH(CH_3)_2]CH_2OH$	$-CH_3$	−H	−H
CN128	$-CH(C_6H_5)CH_2OH$	$-CH_3$	−H	−H

Table 2-1 continued

HPOs	pK_{a1} [a]	pK_{a2} [b]	pK_{a3} [c]	Log K_1	Log β_2	Log β_3	pFe^{3+}
CP20	9.83	3.61		15.01	27.39	37.35	20.59
CP28	9.70	3.69		14.49	26.46	35.96	19.59
CP38	9.88	3.09	3.94	14.62	26.65	36.36	19.46
CP40	9.72	3.39		14.61	26.67	36.34	19.92
CP70	9.33	3.71		14.09	25.80	35.25	19.98
CP102	9.80	3.37		14.69	26.75	36.35	19.69
CP110	9.81	3.26	4.46	14.78	26.79	36.40	19.71
CP111	9.90	3.59	5.84	14.58	26.60	36.03	19.06
CP352	8.56	2.74		13.02	24.03	32.95	19.92
CP359	9.20	2.99		14.26	26.15	35.94	21.06
CP364	9.19	3.06		14.28	26.20	36.01	21.16
CP365	8.93	3.15		14.15	25.68	34.97	20.88
CP366	9.06	3.02		13.82	25.32	34.84	20.35
CP370	9.14	3.00		14.29	26.09	35.66	20.95
CP372	9.15	2.97		14.29	26.06	35.56	20.82
CP374	9.13	3.48		14.41	26.13	35.62	20.94
CP375	9.65	3.29		14.50	26.49	36.27	20.05
CP414	6.96	2.43	10.68	11.59	21.48	30.03	20.16
CP417	6.71	2.03	10.47	10.98	20.54	28.78	19.07
CP510	8.03	2.39		12.20	22.63	31.34	19.71
CP511	6.71	2.42		12.03	22.38	31.26	21.55
CP529	6.43	2.08		11.45	21.53	30.25	20.65
CP545	6.27	1.97		11.32	21.19	29.48	19.92
CP616	10.27	3.56		15.17	27.39	36.83	18.78
CP751	10.32	3.37		15.51	28.12	37.93	19.71
YMF1	8.26	2.65		13.05	24.02	33.01	20.80
YMF3	9.95	4.21		15.06	26.96	35.89	18.80
YMF4	10.51	5.78		16.52	29.33	38.66	19.88
YMF6[d]	10.63	1.06	6.54	14.08	26.16	36.69	17.54
YMF13[d]	11.10	1.42	6.63	14.44	26.97	37.71	17.16
YMMO1	9.90	3.36		14.80	26.81	36.08	19.13
CN116	9.86	3.35		14.82	27.05	36.82	19.98
CN128	9.90	3.26		14.78	27.00	36.79	19.83

[a] Hydroxyl group, [b] carbonyl group, [c] third functional group, the assignments were made using the QM method reported in Chapter 3; [d] the stability constants for Fe^{3+} were determined approximately because the third deprotonated functional group influences the spectra significantly (see Appendix A).

The UV/Vis spectra and corresponding speciation plots for 8-hydroxyquinoline-2-carboxylic acid (8HQC, a tridentate) are presented in Figure 2-4. For pK_a titration, three isosbestic points occur in the spectral series: one at λ 393 nm, corresponding to the formation of $[LH_2]$, the second one at λ 355 nm, corresponding to the formation of $[LH]^-$ and the third one at λ 331 nm, corresponding to the formation of $[L]^{2-}$. $[LH_3]^+$ dominates in solution from pH 0 to pH 0.3. $[LH_2]$ dominates in solution from pH 1 to pH 3. $[LH]^-$ dominates in solution from pH 6 to pH 8. $[L]^{2-}$ dominates in solution from pH 11 to pH 13. The three pK_a values (9.66, 3.91 and 0.62) were determined from curve fitting analysis of the spectra. For Fe^{3+} affinity titration, with a 10 : 1 molar ratio of L : Fe^{3+}, one isosbestic points occur at λ 836 nm when $[Fe^{3+}L_2]^-$ forms. $[Fe^{3+}L_1]^+$ dominates in solution from pH 0.5 to pH 1.5. $[Fe^{3+}L_2]^-$ dominates in solution from pH 2 through to pH 8. The log K_1 and log β_2 values (14.14 and 26.46, respectively) were determined from curve fitting analysis of the spectra. pFe^{3+} (17.73) was determined through export of the stability constants reported above into HySS.

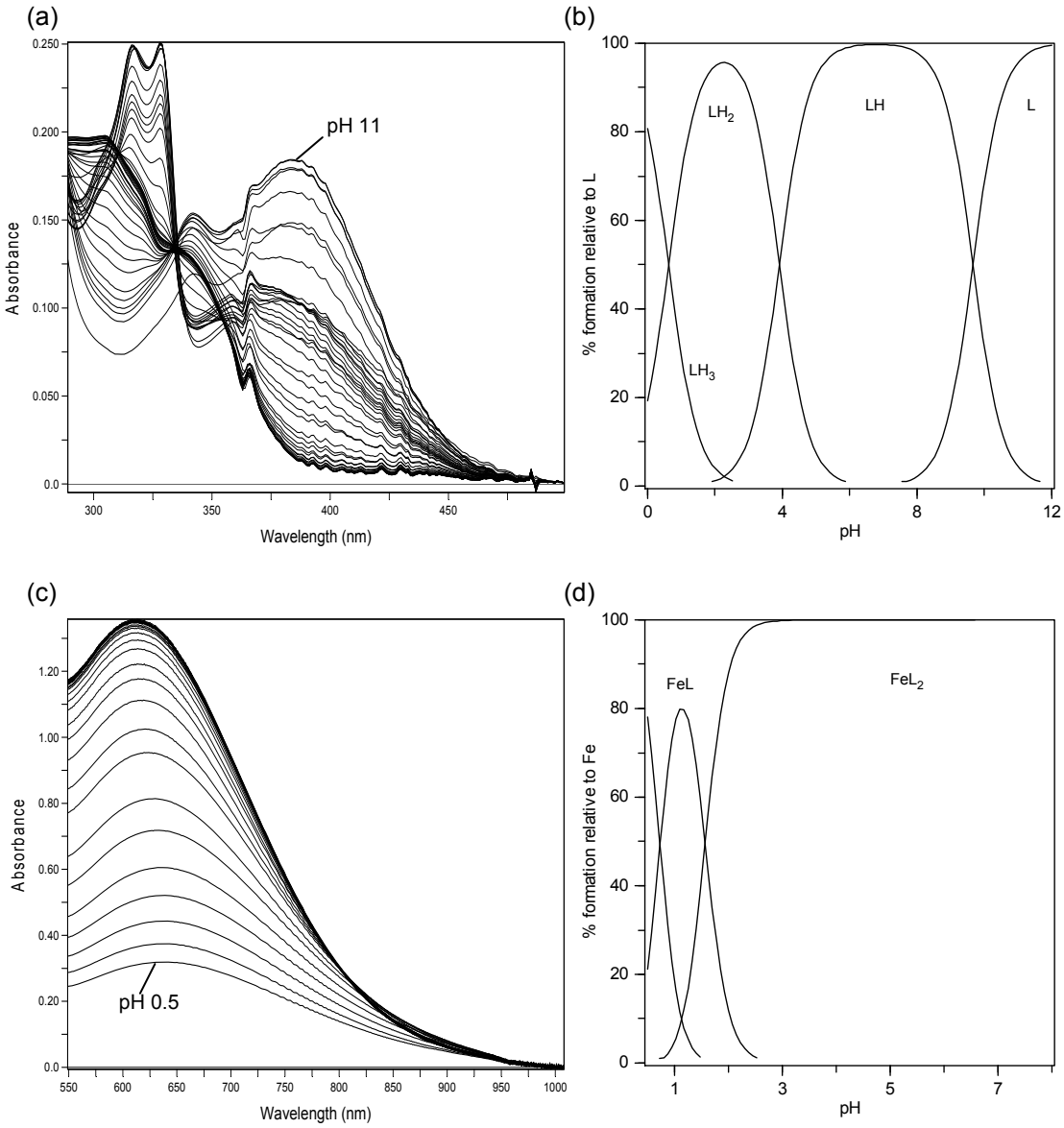
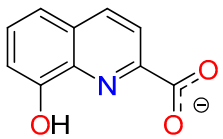


Figure 2-4 UV/Vis spectra and corresponding speciation plots for 8HQC.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 20.5 μM over the pH range 0.5-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 1747.7 μM and [Fe^{3+}] = 178.5 μM over the pH range 0.5-8.



2.4 Discussion

In general, different substituents around the HPO aromatic ring can affect the pK_a values of the two 3,4-oxygen atoms significantly, thereby influencing their affinity for Fe^{3+}. If methyl or alkyl groups are attached to the HPO ring, their electron donating effects can generally increase the pK_a values of the two oxygen atoms. If amido groups are attached to the HPO ring, their electron withdrawing effects and potential formation of intramolecular hydrogen bonding can generally decrease the pK_a values of the two oxygen atoms. If hydroxymethyl groups are attached to the HPO ring, they can result in the same effects as amido groups to decrease the pK_a values of the two oxygen atoms. However, the influence of hydroxymethyl groups is generally weaker than that of amido groups. If fluorine atoms are attached to the HPO ring, their strong electron withdrawing effects can generally decrease the pK_a values of the two oxygen atoms.

Based on these general observations, it is clear that a high electron density on the ligating oxygen atoms leads to both an increase in the pK_a values of the ligand and its affinity for Fe^{3+} in the absence of competitors.

The experimental errors incurred in pK_a determination are mainly contributed by deficiencies in the pH meter employed, for example device ageing, electrode instability, and CO_2 disturbance (Kong, 2008). Such deficiencies are estimated to give rise to systematic errors of the order of ± 0.1 log unit. The experimental errors incurred in (Fe^{3+}) log K_1, log β_2 and log β_3 determinations for HPOs arise mainly as a result of sample concentration errors and errors in the measured hydroxyl and carbonyl pK_a values (particularly the former) (Kong, 2008). Such errors are estimated to be of the order of ± 0.15 log unit, ± 0.31 log unit and ± 0.41 log unit, respectively (assuming errors on the determined pK_a values of ± 0.1 log unit and a sample concentration error of ± 10 %).

The 10 : 1 molar ratio of L : Fe^{3+} condition was adopted to improve the accuracy of the log K_1 experimental determination. The discussion relating to this modification of the standard condition and the general reliability of the experimental log K_1 values is presented in Chapter 5.

Using these titration techniques and the QM calculations (Chapter 3) to assign pK_a values, many previously determined and published Fe^{3+} affinity constants were corrected. These corrections are reported in Chapters 3 and 5.

Chapter 3 Prediction of HPO Hydroxyl pK_a Values Using QSPR and Quantum Mechanical Static Calculations with Implicit Solvent Models

3.1 Introduction

pK_a is one of key physicochemical properties for Fe^{3+} chelators, strongly associated with Fe^{3+} affinity (log β and pFe^{3+}) (Hider and Hall, 1991). In order to predict pK_a values rapidly and reliably, this was made by means of QSPR and QM static calculations using thermodynamic cycles or regression models with implicit solvent models. The assignments of HPO pK_a values reported in Chapter 2 were assisted using the QM static calculations described here for the energy comparison of their deprotonated tautomers. The main results of this chapter are published in Chen *et al.* (2012a).

3.2 Methods

3.2.1 Development of a prediction method based on QM static calculations

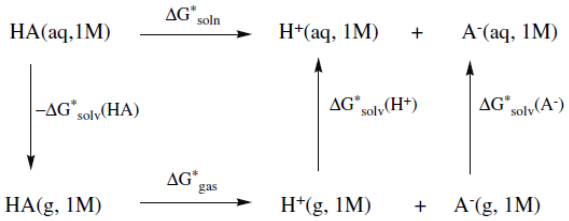
One way of predicting pK_a values is to calculate the free energy in solution (ΔG^*_{soln}) of the protonation process from thermodynamic cycles (Schemes 1 and 2) using the relationship:

$$pK_a = \Delta G^*_{soln} [RT \ln(10)]^{-1}$$

where R is the gas constant (1.99 cal $K^{-1}mol^{-1}$), T is the absolute temperature and K_a is the molar equilibrium constant of the acid.

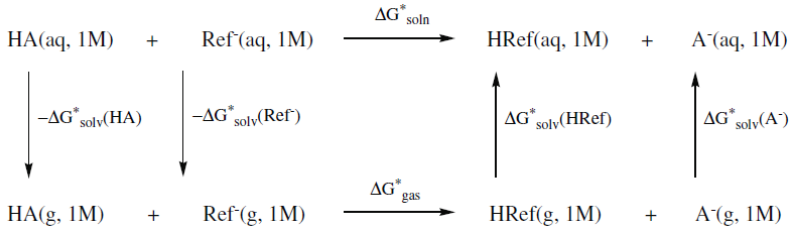
Scheme 3-1 presents the simplest thermodynamic cycle to predict pK_a values and can also be referred to as the direct or absolute method (Ho and Coote, 2010). This method includes accurate gas-phase acidity measurements, from either experimental data or computed by high level *ab initio* or density functional theory (DFT) methods, and solvation free energies gained from different solvent models. In addition, this scheme requires the experimental solvation free energy of the proton, $\Delta G^*_{solv}(H^+)$ = -265.9 kcal mol^{-1} (Ho and Coote, 2010). In general, the gas-phase acidity constant computed *via* high level *ab initio* or DFT methods is believed to be reasonably accurate (with calculated and experimental values typically within ± 2 kcal mol^{-1}) (Foresman and Frisch, 1996). However, there are no gas phase acidity data available for 3-hydroxypyridin-4-ones (HPOs) and only their pK_a values in 0.1M KCl at 25˚C are

available. Dielectric continuum solvent models, treating the solvent as a continuum of uniform dielectric constant (ε) and placing the solute into a cavity within the solvent (Foresman and Frisch, 1996), are thought to have significant errors associated with the calculation of solvation free energies, particularly if the solutes of interest are ionic species or involve hydrogen bond acceptor and/or donor groups (Ho and Coote, 2010). Again, there are no suitable solvation free energy data for HPOs.



Scheme 3-1 pK_a prediction *via* the direct/absolute method (Ho and Coote, 2010).

In order to cancel the systematic errors, mainly related to the solvation free energies, the proton exchange scheme (Scheme 3-2) was introduced. This scheme includes a reference acid (HRef) and although the accuracy of the method is sensitive to the choice of HRef, the proton exchange scheme normally produces better pK_a predictions than the direct method (Ho and Coote, 2010).



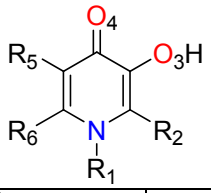
Scheme 3-2 pK_a prediction *via* the proton exchange scheme, HRef: a reference acid (Ho and Coote, 2009).

In the work reported here, Gaussian 09 (Frisch et al., 2009) was utilised to calculate optimised geometries, without symmetry constraints and with force constants calculated at each step and their corresponding free energies obtained from Gaussian 09 outputs. Unless otherwise indicated, all the solvation free energies were derived through direct optimisation in solvent models. The optimised structures were verified as corresponding to a local minimum – a stationary point on the potential energy surface – by checking that there were no imaginary frequencies in the output. In order to reach a global (or nearly global) minimum structure, before running Gaussian 09 optimisation, a conformation search function at the semi-empirical AM1

level in HyperChem Release 8 (http://www.hyper.com/) was carried out for all the molecules studied.

Initially, the methods were developed using the experimentally determined hydroxyl pK_a values of fifteen HPOs (Table 3-1). In order to compare the accuracy of the different methods, the correlation coefficient (r) for the regression between predicted and experimental values and their mean and standard deviation of absolute deviations ($|M|$ and $|S|$) were used.

Table 3-1 Structures and predicted (experimental) hydroxyl pK_a values of sixty-five 3-hydroxypyridin-4-ones (HPOs).



HPOs	R_1	R_2	R_5	R_6
CP20	$-CH_3$	$-CH_3$	−H	−H
CP21	$-C_2H_5$	$-CH_3$	−H	−H
CP23	$-CH(CH_3)_2$	$-CH_3$	−H	−H
CP60	$-CH_3$	−H	−H	−H
CP61	$-C_2H_5$	−H	−H	−H
CP68	$-CH_3$	$-CH_3$	−H	$-CH_3$
CP69	$-CH_3$	−H	−H	$-CH_3$
CP90	−H	−H	−H	−H
CP93	$-CH_3$	$-C_2H_5$	−H	−H
CP94	$-C_2H_5$	$-C_2H_5$	−H	−H
CP99	−H	$-C_2H_5$	−H	−H
CP502	$-CH_3$	$-CONHCH_3$	−H	$-CH_3$
CP511	−H	$-CONHCH_3$	−H	$-CH_3$
CP40	$-C_2H_4OH$	$-CH_3$	−H	−H
CP102	$-C_2H_4OH$	$-C_2H_5$	−H	−H
CP359	$-C_2H_5$	$-CH_2OH$	−H	−H
CP360	$-CH_3$	$-CH_2OH$	−H	$-CH_3$
CP361	$-CH_3$	$-CHCH_3OH$	−H	$-CH_3$
CP364	$-CH_3$	$-CH_2OH$	−H	−H
CP365	$-C_2H_5$	$-CHCH_3OH$	−H	−H
CP366	$-C_3H_6OH$	$-CH_2OH$	−H	−H
CP369	$-C_2H_5$	$-CH_2OH$	−H	$-CH_3$
CP372	$-C_3H_6OH$	$-CHCH_3OH$	−H	−H
CP374	$-CH_3$	$-CHC_2H_5OH$	−H	$-CH_3$
CP508	$-CH_3$	$-CON(CH_3)_2$	−H	$-CH_3$
CP509	−H	$-CON(CH_3)_2$	−H	$-CH_3$
CP510	$-CH_3$	−H	−H	$-CONHCH_3$
CP513	$-CH_3$	$-CONHC_2H_5$	−H	$-CH_3$
CP514	$-C_2H_5$	$-CONHCH_3$	−H	$-CH_3$
CP528	−H	$-CONH_2$	−H	−H
CP545	−H	$-CONHC_2H_4OH$	−H	−H
YMF1	−H	$-CHCF_3OH$	−H	−H
YMF5	−H	−H	−F	−H
YMF7	$-CH_3$	−F	−H	−H
YMF8	$-C_2H_5$	−F	−H	−H
YMF10	−H	−F	$-CONHCH(CH_3)CONHCH_3$	−H
YMF13	−H	−F	$-CH_3$	−H
YMF14	$-CH_3$	−F	$-CH_3$	−H
YMF15	$-CH_3$	−H	−F	−H
YMF16	$-C_2H_5$	−H	−F	−H
YMF17	$-CH_3$	$-CH_3$	−F	−H
YMF18	−H	−F	−H	$-CH_3$
YMF19	−H	−F	$-CH_3$	$-CH_3$
YMF21	−H	−H	−F	$-CH_3$
YMF22	−H	$-CH_3$	−F	−H
YMF24	$-C_3H_7$	−F	−H	−H
YMF25	$-C_3H_7$	−H	−F	−H
YMF26	$-CH(CH_3)_2$	−H	−F	−H
YMF29	−H	−F	$-CONHCH_3$	−H
YMF31	−H	−F	$-CONH_2$	−H
YMF32	−H	−F	$-CON(CH_3)_2$	−H
YMF33	$-C_4H_9$	−F	−H	−H
CP38	$-C_2H_4COOH$	$-CH_3$	−H	−H
CP110	$-C_2H_4COOH$	$-C_2H_5$	−H	−H
CP111	$-C_3H_6COOH$	$-C_2H_5$	−H	−H
CP352	$-C_2H_4OH$	$-CH(C_6H_5)OH$	−H	$-CH_3$
CP362	$-CH_3$	$-CH_2OCH_3$	−H	$-CH_3$
CP363	$-CH_3$	$-CH(CH_3)OCH_3$	−H	$-CH_3$
CP414	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_3$
CP417	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_2OH$
YMF3	−H	−F	−F	−F
YMF4	−H	−F	−H	−F
YMF6	−H	−F	−H	−H
CP751	$-CH_3$	$-CH_3$	$-CH_3$	−H
YMF11	−H	$-CONHCH(CH_3)CONHCH_3$	−F	−H

Table 3-1 continued

HPOs	$\Delta G^*_{calculated}$ ($G^*_{calculated}$ for LH, L^-) (Hartree)[i]	Hydroxyl pK_a
CP20[a]	0.4598 (-477.2887, -476.8289)	9.76 (9.76)[b]
CP21[a]	0.4609 (-516.5787, -516.1179)	10.04 (9.82)[b]
CP23[a]	0.4610 (-555.8663, -555.4053)	10.07 (9.90)[b]
CP60[a]	0.4559 (-437.9967, -437.5408)	8.81 (8.79)[b]
CP61[a]	0.4573 (-477.2883, -476.8311)	9.15 (8.87)[b]
CP68[a]	0.4615 (-516.5789, -516.1174)	10.19 (10.13)[b]
CP69[a]	0.4575 (-477.2890, -476.8316)	9.20 (9.12)[b]
CP90[a]	0.4567 (-398.7130, -398.2563)	9.01 (9.01)[b]
CP93[a]	0.4598 (-516.5744, -516.1146)	9.77 (9.85)[b]
CP94[a]	0.4604 (-555.8647, -555.4043)	9.92 (9.86)[b]
CP99[a]	0.4595 (-477.2971, -476.8376)	9.70 (9.81)[b]
CP502[a]	0.4554 (-685.2680, -684.8126)	8.69 (8.44)[b]
CP511[a]	0.4471 (-646.0009, -645.5538)	6.66 (6.66)[b]
CP40[a]	0.4595 (-591.7952, -591.3356)	9.70 (9.72)[c]
CP102[a]	0.4600 (-631.0816, -630.6216)	9.81 (9.76)[c]
CP359	0.4570 (-591.7920, -591.3350)	9.08 (9.05)[d]
CP360	0.4572 (-591.7924, -591.3352)	9.14 (9.42)[d]
CP361	0.4539 (-631.0809, -630.6270)	8.33 (8.97)[d]
CP364	0.4567 (-552.5020, -552.0453)	9.02 (9.11)[d]
CP365	0.4537 (-631.0811, -630.6275)	8.27 (8.74)[d]
CP366	0.4571 (-706.2979, -705.8408)	9.12 (9.11)[d]
CP369	0.4567 (-631.0810, -630.6243)	9.00 (9.38)[d]
CP372	0.4536 (-745.5875, -745.1339)	8.24 (8.69)[d]
CP374	0.4540 (-670.3692, -669.9152)	8.36 (8.98)[d]
CP508	0.4546 (-724.5470, -724.0923)	8.51 (8.14)[e]
CP509	0.4532 (-685.2715, -684.8183)	8.16 (8.20)[e]
CP510	0.4537 (-645.9762, -645.5224)	8.29 (8.05)[e]
CP513	0.4555 (-724.5588, -724.1033)	8.72 (8.45)[e]
CP514	0.4554 (-724.5566, -724.1012)	8.69 (8.41)[e]
CP528	0.4459 (-606.7032, -606.2574)	6.36 (6.38)[e]
CP545	0.4448 (-721.2113, -720.7665)	6.09 (6.27)[e]
YMF1	0.4529 (-850.2826, -849.8297)	8.08 (8.19)[f]
YMF5	0.4537 (-497.9583, -497.5046)	8.27 (8.23)[f]
YMF7	0.4525 (-537.2402, -536.7877)	7.98 (7.72)[f]
YMF8	0.4531 (-576.5310, -576.0779)	8.13 (7.72)[f]
YMF10	0.4609 (-952.7847, -952.3238)	10.04 (9.97)[f]
YMF13	0.4642 (-536.8056, -536.3414)	10.85 (11.28)[f]
YMF14	0.4543 (-576.5339, -576.0797)	8.42 (8.35)[f]
YMF15	0.4534 (-537.2414, -536.7880)	8.21 (8.18)[f]
YMF16	0.4539 (-576.5335, -576.0796)	8.33 (8.26)[f]
YMF17	0.4571 (-576.5342, -576.0771)	9.12 (9.08)[f]
YMF18	0.4639 (-536.8085, -536.3446)	10.77 (11.08)[f]
YMF19	0.4658 (-576.0979, -575.6320)	11.25 (11.56)[f]
YMF21	0.4547 (-537.2553, -536.8006)	8.52 (8.69)[f]
YMF22	0.4559 (-537.2553, -536.7994)	8.82 (8.93)[f]
YMF24	0.4531 (-615.8196, -615.3665)	8.12 (7.89)[f]
YMF25	0.4538 (-615.8220, -615.3682)	8.30 (8.12)[f]
YMF26	0.4541 (-615.8226, -615.3685)	8.38 (8.12)[f]
YMF29	0.4603 (-705.5090, -705.0487)	9.90 (9.95)[f]
YMF31	0.4600 (-666.2261, -665.7661)	9.81 (9.96)[f]
YMF32	0.4612 (-744.7795, -744.3183)	10.11 (10.26)[f]
YMF33	0.4534 (-655.1086, -654.6553)	8.20 (7.79)[f]
CP38	0.4610 (-704.7183, -704.2573)	10.08 (9.88)[g]
CP110	0.4613 (-744.0059, -743.5446)	10.14 (9.81)[g]
CP111	0.4612 (-783.2938, -782.8326)	10.12 (9.90)[g]
CP352	0.4574 (-937.2790, -936.8216)	9.19 (8.56)[g]
CP362	0.4591 (-631.0699, -630.6107)	9.61 (9.31)[g]
CP363	0.4600 (-670.3575, -669.8975)	9.88 (9.58)[g]
CP414	0.4477 (-767.6005, -767.1527)	6.81 (6.96)[g]
CP417	0.4461 (-842.8134, -842.3673)	6.43 (6.71)[g]
YMF3	0.4589 (-696.0163, -695.5574)	9.55 (9.95)[g]
YMF4	0.4619 (-596.7734, -596.3114)	10.30 (10.32)[g]
YMF6	0.4629 (-497.5130, -497.0500)	10.54 (10.63)[g]
CP751	0.4615 (-516.5825, -516.1210)	10.18 (10.32)[g]
YMF11	0.4427 (-953.2248, -952.7821)	5.58 (5.15)[h]

[a] the first fifteen HPOs were used to test different existing pK_a prediction methods; experimental pK_a values taken from [b] Kong (2008); [c] Rai *et al.* (1998); [d] Liu *et al.* (1999); [e] Piyamongkol *et al.* (2010); [f] Ma *et al.* (2012); [g] this study obtained from Chapter 2; [h] Kong (Chen *et al.*, 2012a); [i] $\Delta G^*_{calculated}$: the calculated free energy differences between LH and L^- using B3LYP/6-31+G(d)/CPCM(Pauling radii, water as solvent) ($G^*_{calculated}$: the calculated free energy); 1 Hartree = 627.5095 kcal mol^{-1}.

Test 1: the direct method was applied using model chemistry B3LYP/6-31+G(d), a DFT method widely used for ionic species (Foresman and Frisch, 1996) and CPCM solvent model in water with default cavity radii, UFF radii, denoted as B3LYP/6-31+G(d)/CPCM(UFF radii, water as solvent). The pK_a predictions achieved were reasonably accurate (giving r^2 = 0.90, |M| = 5.50 and |S| = 1.76, Figure 3-1).

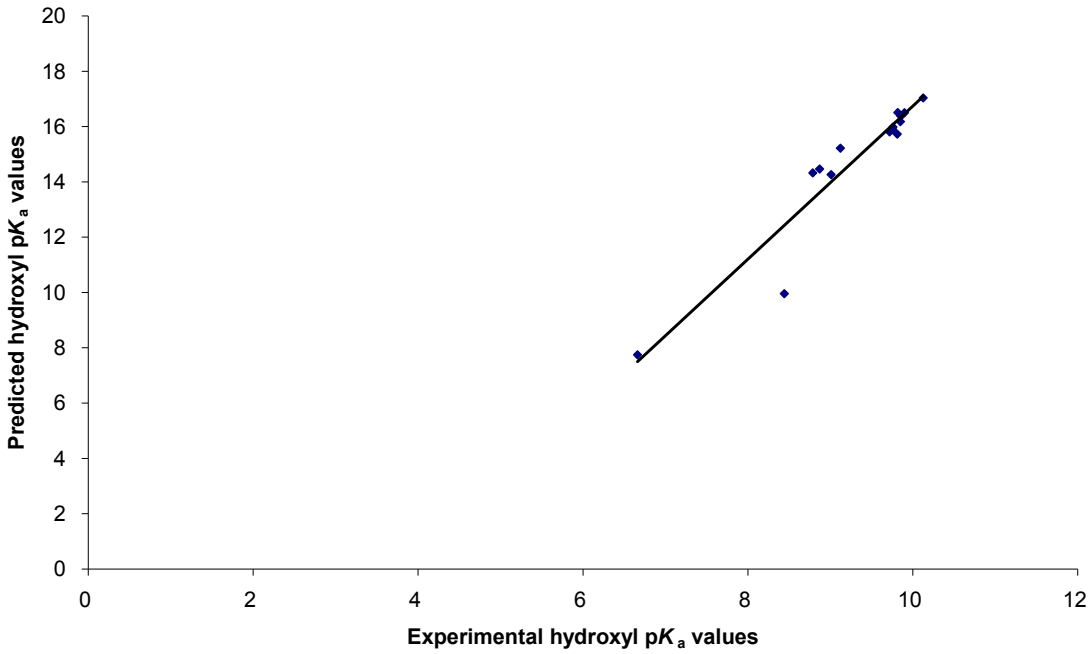


Figure 3-1 Result of Test 1, hydroxyl pK_a values predicted following the procedure presented in Scheme 3-1 with free energies computed using a B3LYP/6-31+G(d) basis set and the CPCM default model to treat solvent effects.

Test 2: the proton exchange scheme (using CP90 as a reference acid) was implemented using the same model chemistry and solvent model as employed in Test 1. The pK_a predictions achieved in this scheme were rather better than those obtained using the direct method (giving r^2 = 0.90, |M| = 1.30 and |S| = 1.17, Figure 3-2).

The tests using the proton exchange scheme *via* higher level model chemistries for solvation free energies were also explored and the results are indicated in Table 3-2. In these tests, the solvation free energies were calculated from single-point energy calculations, rather than from optimisations. From Tests 3 and 4, it was established that the changes in the basis set or DFT functional do not give significantly better predictions than those achieved in Test 2. From Tests 5, 6 and 7, compound methods (CBS-X or Gaussian-X) (Foresman and Frisch, 1996) were found to be incapable of calculating pK_a values for some or all of the fifteen HPOs

owing to computer resource limitations. Moreover, these methods require much more computing time when compared with the DFT methods.

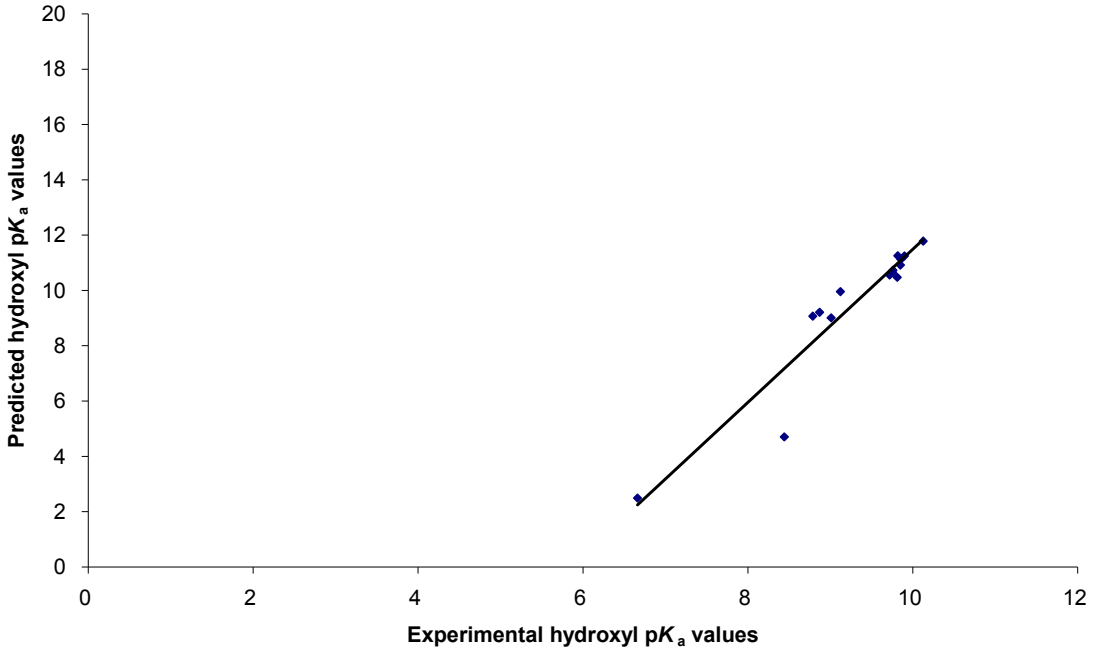


Figure 3-2 Result of Test 2, the proton exchange scheme (using CP90 as a reference acid) was applied using the same model chemistry as Test 1.

Table 3-2 Results of tests using the proton exchange scheme *via* higher level model chemistries for solvation free energies.

Tests	Model Chemistry	Results
3[a]	B3LYP/6-311++G(3df,3pd)/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d)	There are no significant improvements compared to Test 2.
4	B971/6-311++G(3df,3pd)/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d)	There are no significant improvements compared to Test 2.
5	CBS-4M/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d)	Some of the compounds could not be modelled using the Windows Gaussian 09 system owing to computer resource limitations.
6	CBS-4Q/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d)	Some of the compounds could not be modelled using the Windows Gaussian 09 system owing to computer resource limitations.
7	Gaussian-4/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d)	Some of the compounds could not be modelled using the Windows Gaussian 09 system owing to computer resource limitations.

[a] B3LYP/6-311++G(3df,3pd)/CPCM(UFF radii, water as solvent)//B3LYP/6-31+G(d) means that the single-point energy calculation at B3LYP/6-311++G(3df,3pd)/CPCM(UFF radii, water as solvent) based on the optimised structure at B3LYP/6-31+G(d) in vacuum.

In order to simplify the calculations, only the optimised structures in solvent models were calculated. The calculated free energy differences between the neutral and deprotonated structures ($\Delta G^*_{calculated}$) of the fifteen HPOs were plotted against their experimental hydroxyl pK_a values to search for correlations. The r^2 values of the regressions were taken to evaluate which model chemistry gave the best performance. For deprotonated structures, there are two ways to obtain the calculated free energy. One is to calculate a single-point energy for the deprotonated form of the optimised neutral structures and the other is to further optimise the structure. The different model chemistries explored differ mainly in the nature of the solvent models used

(CPCM, IEF-PCM, I-PCM, SCI-PCM or SMD) and the radii employed in the CPCM models (UFF, UAHF, UAKS, Bonding, or Pauling).

The model chemistry that provided the best results was found to be B3LYP/6-31+G(d)/CPCM(Pauling radii, water as solvent), with r^2 = 0.98 (Figure 3-3). For this method, all the neutral and deprotonated forms required optimisation. B3LYP/3-21G (a lower basis set), CAM-B3LYP/6-31+G(d) (a modified functional which is suitable for long range calculations) and PM6 (a semi-empirical method) with the same solvent model were also tested. However, they did not produce superior outcomes with respect to either accuracy or computing time. (Any further calculations including, for example, explicit solvent were not performed since the additional computing time required could not be justified.)

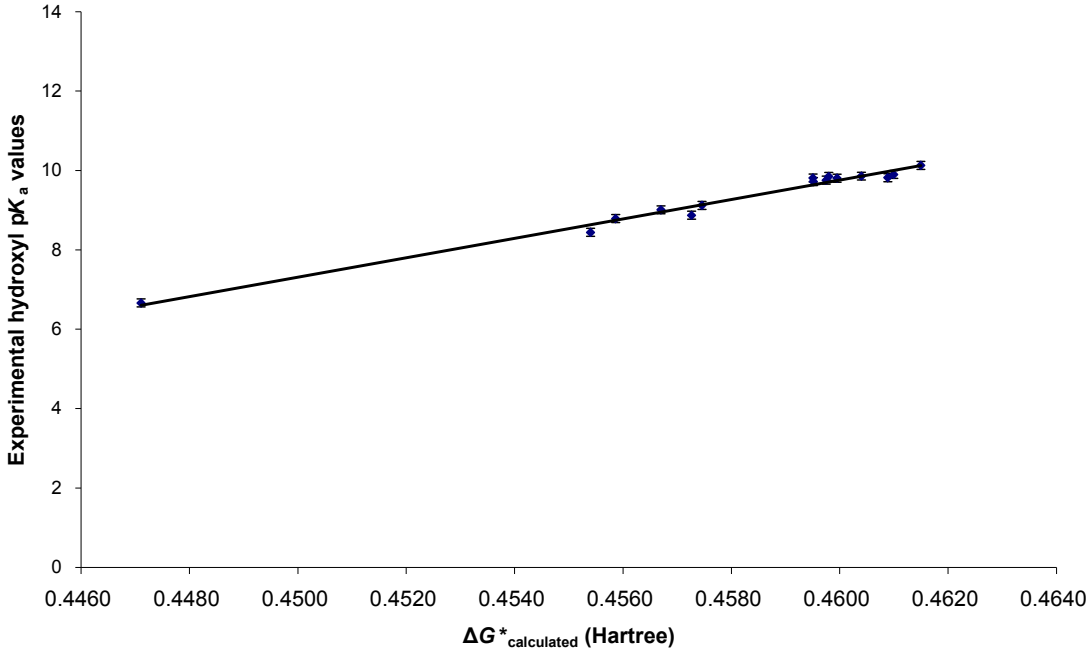


Figure 3-3 Correlation of experimental hydroxyl pK_a values[a] against the calculated free energy differences ($\Delta G^*_{calculated}$) for the starting fifteen HPOs using B3LYP/6-31+G(d)/CPCM (Pauling radii, water as solvent).

[a] The errors related to the experimental hydroxyl pK_a determination for the typical HPOs are estimated to be of the order of ± 0.1 (see discussion in Chapter 2). 1 Hartree = 627.5095 kcal mol^{-1}.

3.2.2 Deprotonation sequence determination

With the aid of QM static calculations, the assignment of deprotonation sequence can be decided by comparing the potential energies of the possible tautomers. If their potential energies differ by more than 0.6 kcal mol^{-1}, the tautomer with the lowest potential energy can be considered to be the major deprotonated form in solution. (Calculations involving Boltzmann

averaging over multiple low energy conformers were not performed since the additional computing time required could not be justified.)

3.2.3 QSPR pK_a prediction methods

The hydroxyl pK_a values of the sixty-five HPOs (Table 3-1) were also predicted using ACD/pK_a DB 12.0 (http://www.acdlabs.com/) and Marvin 5.4.0.1 (http://www.chemaxon.com/). The ACD/pK_a DB 12.0 method is centred on linear free energy relationships, using the empirical relations of Hammett and Taft (Liao and Nicklaus, 2009). The Marvin 5.4.0.1 method is centred on quantitative structure-property relationships, associating calculated structural descriptors with experimental pK_a values (Liao and Nicklaus, 2009).

3.3 Results

3.3.1 pK_a predictions using ACD and Marvin

The ACD/pK_a DB 12.0 method fails to provide accurate hydroxyl pK_a values for any of the sixty-five HPOs included in this study (Figure 3-4). In particular, values for the 2-amido-, 2-(piperidin-1-ylmethyl)-, 2-fluoro and 2,6-difluoro-HPO subclasses are poorly predicted. The Marvin 5.4.0.1 method also fails to provide accurate predictions (Figure 3-5), the poorest being for the 2-(piperidin-1-ylmethyl)- and 2-fluoro-HPO subclasses.

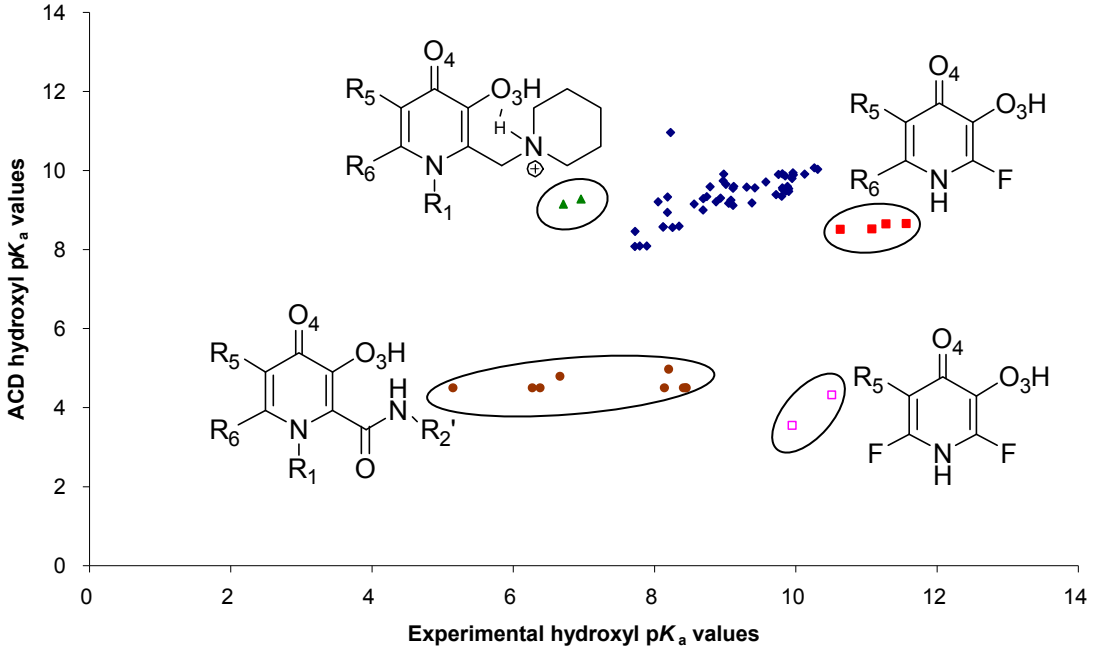


Figure 3-4 ACD predictions for hydroxyl pK_a values. ▲ 2-(piperidin-1-ylmethyl)-HPOs; ■ 2-fluoro-HPOs; ● 2-amido-HPOs; ▫ 2,6-difluoro-HPOs.

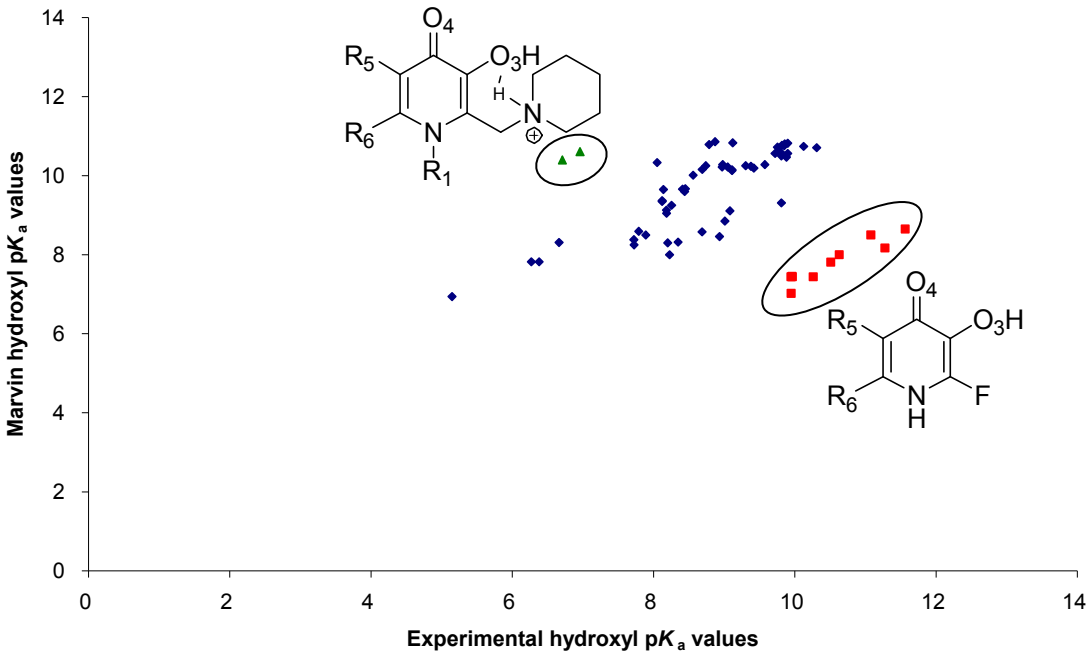


Figure 3-5 Marvin predictions for hydroxyl pK_a values. ▲ 2-(piperidin-1-ylmethyl)-HPOs; ■ 2-fluoro-HPOs.

3.3.2 Developed pK_a prediction model

As discussed above, B3LYP/6-31+G(d)/CPCM(Pauling radii, water as solvent) generated comparatively reliable predictions of hydroxyl pK_a values for a series of fifteen HPOs. The pK_a values for an additional forty-eight HPOs were computed using the same model chemistry to develop the final pK_a prediction model — Predicted pK_a = 245.16 $\Delta G^*_{calculated}$ - 102.95 (r^2 = 0.95, |M| = 0.20 and |S| = 0.16, Figure 3-6; values in Table 3-1). Unlike the ACD/pK_a DB 12.0 and Marvin 5.4.0.1 methods, this model provides a good correlation between the predicted and experimental values.

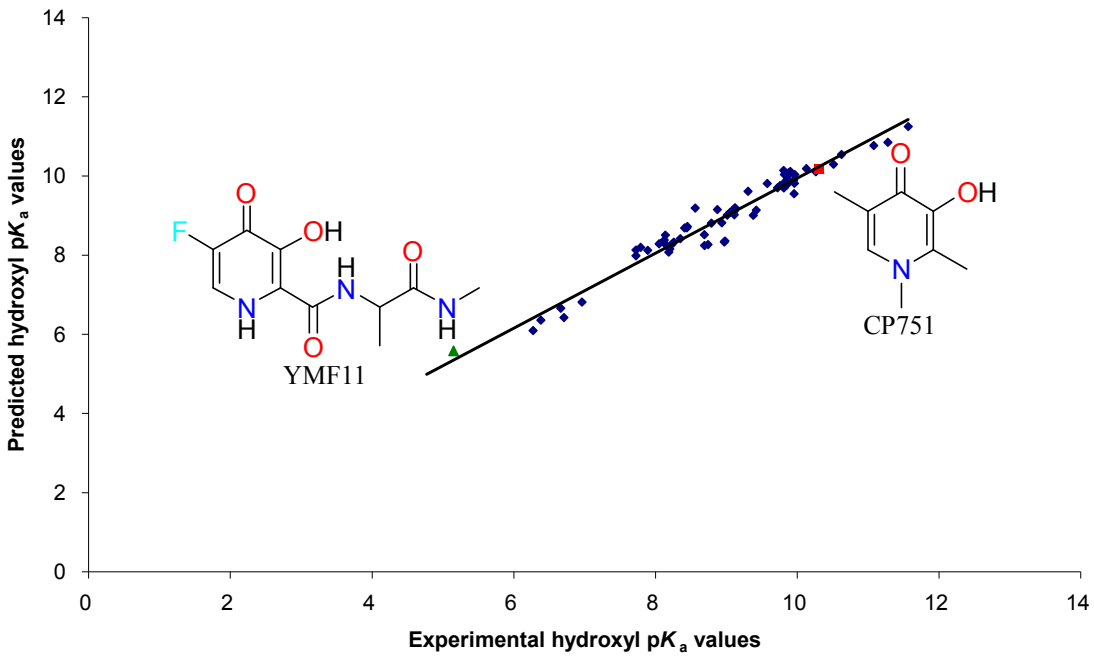


Figure 3-6 Developed pK_a prediction model. ▲ YMF11; ■ CP751.

In order to further investigate the predictive power of this model, two novel HPOs, CP751 and YMF11, were synthesised by Ma (Chen *et al.*, 2012a). The experimental hydroxyl pK_a value of CP751 is within the pK_a training range while that of YMF11 is outside the pK_a training range (Figure 3-6). However, the data points for both molecules fell close to the regression line, indicating reliable predictions. The absolute deviation between the predicted and experimental pK_a values for CP751 (10.18, 10.31) is 0.13 and for YMF11 (5.58, 5.15) is 0.43.

3.4 Discussion

With HPOs possessing more than two pK_a values, the assignment of pK_a values can become complicated. This is true in the cases of 2-(piperidin-1-ylmethyl)-HPOs (e.g. CP417) and some 2-fluoro-HPOs (e.g. YMF6). The deprotonation sequence of CP417 was previously incorrectly assigned (Figure 3-7c) and as a result, the hydroxyl pK_a value was assigned as 10.51 (Dehkordi *et al.*, 2008). However, QM static calculations for the tautomers (Figure 3-8) suggest a different deprotonation sequence (Figure 3-7b) and the hydroxyl pK_a is assigned as 6.71.

The fact that the hydroxyl pK_a value of 2-(piperidin-1-ylmethyl)-HPOs is less than 7 is tentatively attributed to the formation of a strong intramolecular hydrogen bond (Figure 3-9a), by analogy with the hydroxyl pK_a value of 2-amido-HPOs (e.g. CP511, Figure 3-9b), which are also

less than 7. The hydrogen bonding depicted in Figure 3-9b has been confirmed by X-ray diffraction (Piyamongkol *et al.*, 2010).

The pK_a assignments for some 2-fluoro-HPOs are not straightforward. If 2-fluoro-HPOs have a 1-alkyl group, there is no complexity in the assignments, and such compounds behave as typical alkyl HPOs. In contrast, if 2-fluoro-HPOs have a 1-hydrogen group, the assignments become complicated owing to the generation of a third pK_a value (Figure 3-7d). Based on QM static calculations, it is suggested that the NH group deprotonates before the hydroxyl group (Figure 3-7d). This deprotonation results in an electronic effect, leading to the oxygen of the hydroxyl group attracting the proton more tightly. Thus, the hydroxyl pK_a values in this series of compounds are higher than in alkyl HPOs.

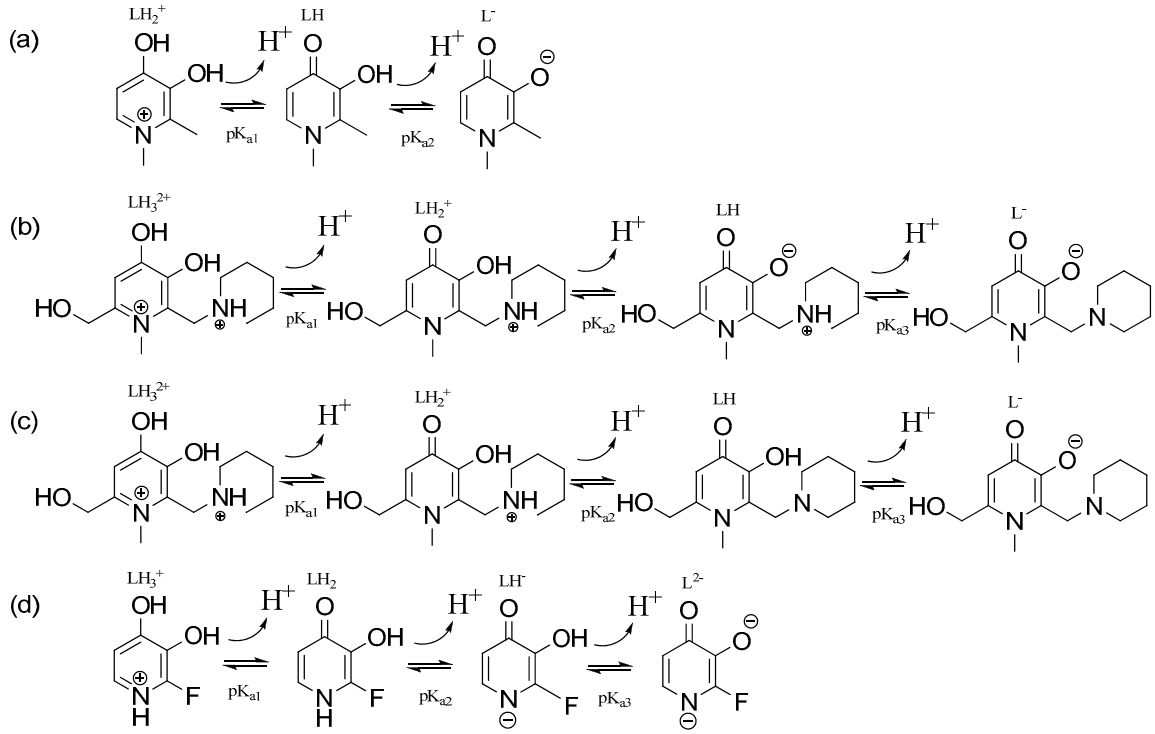


Figure 3-7 (a) Deprotonation sequence of CP20.
(b) Deprotonation sequence of CP417.
(c) Deprotonation sequence of CP417. proposed by Dehkordi *et al.* (2008).
(d) Deprotonation sequence of YMF6.

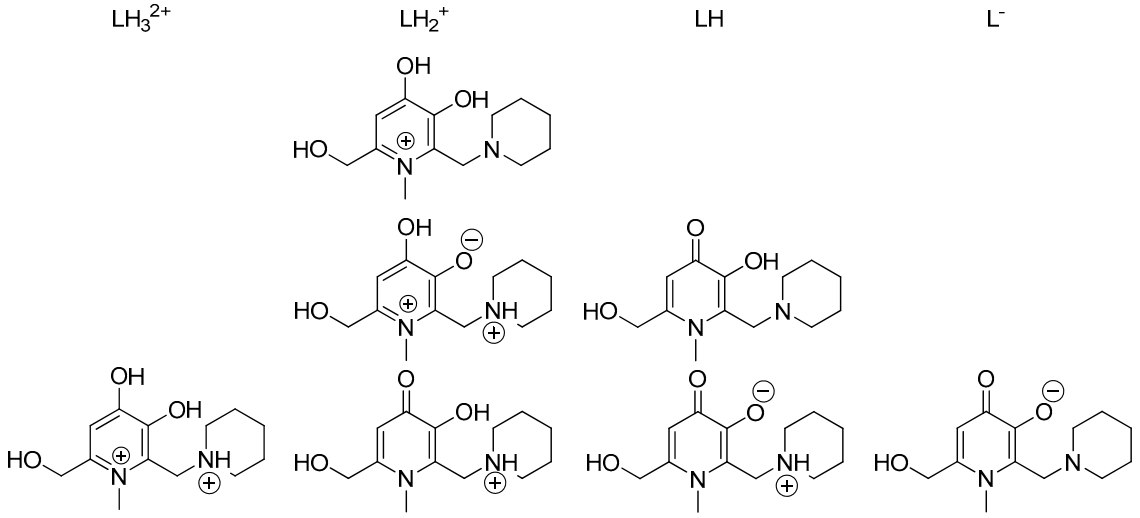


Figure 3-8 CP417 tautomers[a].
[a] The position of a tautomer in a group provides an indication of its energy content, the lower the position, the lower the energy content.

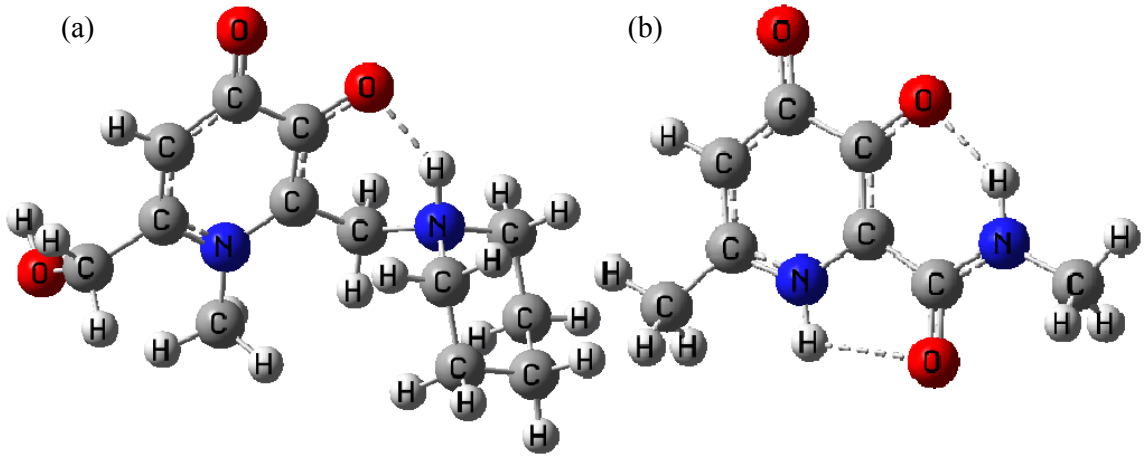


Figure 3-9 (a) intramolecular hydrogen bond for CP417; (b) intramolecular hydrogen bonds for CP511.

This study demonstrates that there are two broad classes of outliers for both the ACD/pK_a DB 12.0 and Marvin 5.4.0.1 programs (Figures 3-4 and 3-5). One is centred on HPOs possessing intramolecular hydrogen bonding with the 3-oxygen atom, such as CP417, a 2-(piperidin-1-ylmethyl)-HPO. The other one is centred on HPOs possessing strong electron withdrawing functional groups, such as YMF6, a 2-fluoro-HPO. It is clear that intramolecular hydrogen bonding and electronic effects are the two main factors which the two commercial prediction programs cannot successfully accommodate unlike QM static calculations based on the approximation of Schrödinger equations.

Chapter 4 Prediction of HPO Hydroxyl pK_a Values Using Quantum Mechanical Molecular Dynamic Simulations with Explicit Solvent

4.1 Introduction

Predictions of three HPOs' hydroxyl pK_a values were made using quantum mechanical molecular dynamic (QM MD) simulations involving explicit solvent and the results obtained were compared with those determined by experiment and obtained using the other theoretical methods described in Chapter 3. In addition to predicting the pK_a values, the QM MD simulations provide valuable additional insights into the atomistic details of the proton transfer mechanism and the hydration structures and dynamics at all stages of the reaction. The hydration properties were further examined by means of radial distribution functions, correlated with the order of their experimentally determined pK_a values. The main results of this chapter are published in Chen *et al.* (2012c).

4.2 Methods

4.2.1 Coordination constrained Car-Parrinello Molecular Dynamics simulations

Constrained Car-Parrinello molecular dynamics (CP-MD) simulations with the CP-MD software (http://www.cpmd.org/) were carried out for three different HPOs (CP20, CP60, and CP751, Figure 4-1) surrounded by 61 H_2O molecules in a periodically repeating cubic unit cell of size 12.4 Å x 12.4 Å x 12.4 Å. The Kohn-Sham equations were solved using the PBE (Perdew *et al.*, 1996) exchange-correlation functional in a plane wave basis set truncated at 25 Ry (340 eV), in conjunction with ultrasoft pseudopotentials (Vanderbilt, 1990). The coupled nuclear and fictitious electronic equations of motion were solved using a timestep of 6 a.u. (0.145 fs), setting the fictitious orbital mass to 800 a.u. (0.439 a.m.u.), and replacing all hydrogens with deuterium. A Nosé-Hoover chain thermostat (Hoover, 1985; Nose, 1984) was applied to maintain the nuclear temperature at 350 K. A slightly raised temperature is necessary in CP-MD simulations with this type of exchange-correlation functional to reproduce structural and dynamical experimental data for liquid water at room temperature (Fernandez-Serra and Artacho, 2004; Schwegler *et al.*, 2004; Sit and Marzari, 2005).

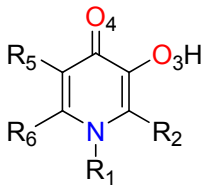


Figure 4-1 Simulated HPO structures. CP20: $R_1 = R_2 = CH_3$ and $R_5 = R_6 = H$; CP60: $R_1 = CH_3$ and $R_2 = R_5 = R_6 = H$; CP751: $R_1 = R_2 = R_5 = CH_3$ and $R_6 = H$.

The number of protons coordinated to the 3-hydroxyl oxygen of HPOs was calculated using the formula:

$$n = \sum_i \frac{1}{\exp[\kappa(r_i - r_c)] + 1} \quad (1)$$

with a cutoff radius of r_c = 2.46 a.u. (1.30 Å) and κ = 5.29 a.u.$^{-1}$ (2.80 Å^{-1}), r_i being the distance between the 3-hydroxyl oxygen and the i^{th} proton. For each fixed proton coordination number n, a trajectory of at least 1.5 ps was calculated and the time-averaged constraint force $< f >$ was determined. Each simulation was run until the running average of the constraint force reached a relative standard deviation of less than 2% during the last 0.7 ps.

The free energy difference between the initial equilibrium state (with the intact 3-hydroxyl group) at the coordination number of n_0, and the final state (at the top of the reaction barrier) at the coordination number of n_f, ΔA, was obtained by thermodynamic integration:

$$\Delta A = A(n_f) - A(n_0) = \int_{n_0}^{n_f} < f'(n) > dn \quad (2)$$

where $<f'(n)>$ is the potential of mean force, the metrically corrected time-averaged constraint force (Sprik and Ciccotti, 1998) $< f >$. The free energy difference was converted to a molar equilibrium constant of acid using the relation (Chandler, 1987; Davies *et al.*, 2002):

$$K_a = \frac{(1-P)^2}{P C_0 V} \quad (3)$$

where

$$P = \frac{\int_0^{R_c} 4\pi r^2 \exp(-\Delta A / kT) dr}{\int_0^{R_f} 4\pi r^2 \exp(-\Delta A / kT) dr} \quad (4)$$

is the probability of finding a proton within the radius R_c, k is the Boltzmann constant, T is the absolute temperature, C_0 is the standard concentration (1 mol dm^{-3}), and V is the volume of the simulation cell. R_c, the critical distance at which the 3-hydroxyl oxygen (O_3) share its proton with H_2O equally, was estimated to be 1.3 Å based on the O_3H radial distribution functions, and R_f was taken as the O_3H distance at the final coordination number n_f, at which O_3 loses its proton completely.

4.3 Results

The potential of mean force and the corresponding free energy profiles of CP-MD simulations at different proton coordination numbers (n) for the three HPOs are presented in Figure 4-2. The simulations were started at thermal equilibrium with the fully intact 3-hydroxyl group, corresponding to n ~ 0.96, at which the mean force is practically zero. When the proton coordination number is decreased, the O_3H bond is stretched, resulting in a sharp rise in the constraint force until it reaches a maximum at n ~ 0.93. When n is decreased further, the mean force reduces towards zero at which the system approaches the transition state for the proton transfer prior to becoming negative when the system enters the product potential. Snapshots for the different stages of the proton transfer reaction are illustrated in Figure 4-3. Simulations carried out from n ~ 0.96 to n ~ 0.3 are sufficient to predict the pK_a values because only the region up to the transition state is relevant for the corresponding free energy barrier. The force curves (Figure 4-2a) show that the highest peak belongs to CP751, followed by CP20 and CP60, the lowest, and this order is reflected in the corresponding free energy profiles (Figure 4-2b). CP751 possesses the largest free energy barrier (ΔA = 36.3 kJ mol^{-1}), followed by CP20 (ΔA = 32.1 kJ mol^{-1}) and CP60 (ΔA = 27.5 kJ mol^{-1}). These free energies were converted to the pK_a values as 10.73 (CP751), 9.36 (CP20), and 9.00 (CP60) (Table 4-1).

The pK_a predictions from the CP-MD simulations are compared to their experimental measurements (Chen *et al.*, 2012a; Kong, 2008) and those obtained *via* the other theoretical methods described in Chapter 3 (Table 4-1). The ACD/pK_a DB 12.0 method yields good predictions for CP751 and CP20 but a significant error for CP60, with the difference between the measured and predicted pK_a values being almost 1 log unit. The Marvin 5.4.0.1 method yields very poor predictions and overestimates the pK_a values for all three HPOs by about 1 log unit. The direct method yields pK_a values that follow the correct trend but with the absolute predicted values overestimated for all three HPOs by greater than 6 log units. The proton

exchange scheme corrects the absolute predicted values derived from the direct method – reducing the average absolute error from 6.25 to 0.99, but still yielding pK_a values of insufficient accuracy. The CP-MD method yields very good predictions with respect to trend and average absolute error (0.34) for all three HPOs. Compared to the other methods, QM static calculations with a regression model developed in Chapter 3 yield the best results.

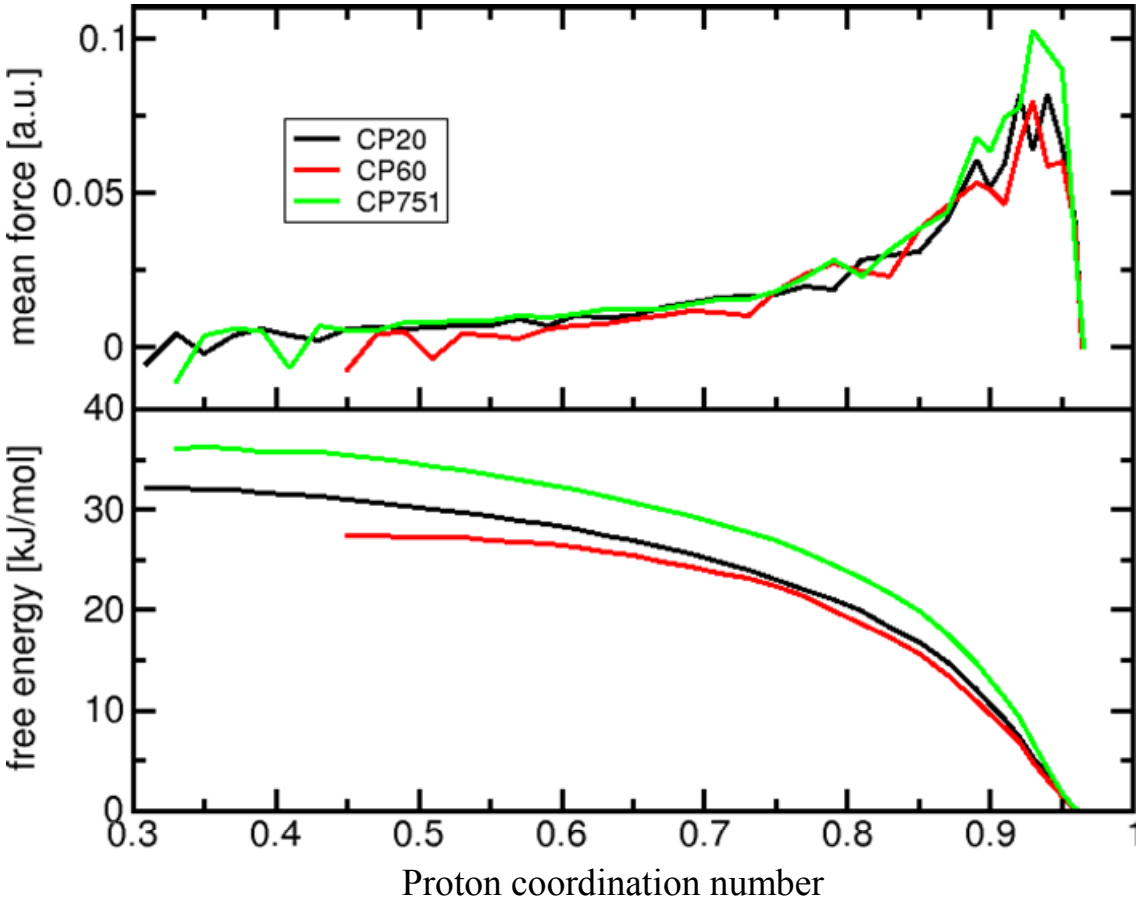
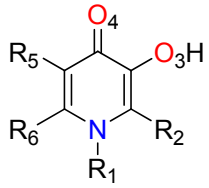


Figure 4-2 Mean force (a) and the corresponding free energy (b) *versus* proton coordination number for the three HPOs[a].

[a] CP20 (black line), CP60 (red line), CP751 (green line).



CP20: $R_1 = R_2 = CH_3$ and $R_5 = R_6 = H$
CP60: $R_1 = CH_3$ and $R_2 = R_5 = R_6 = H$
CP751: $R_1 = R_2 = R_5 = CH_3$ and $R_6 = H$

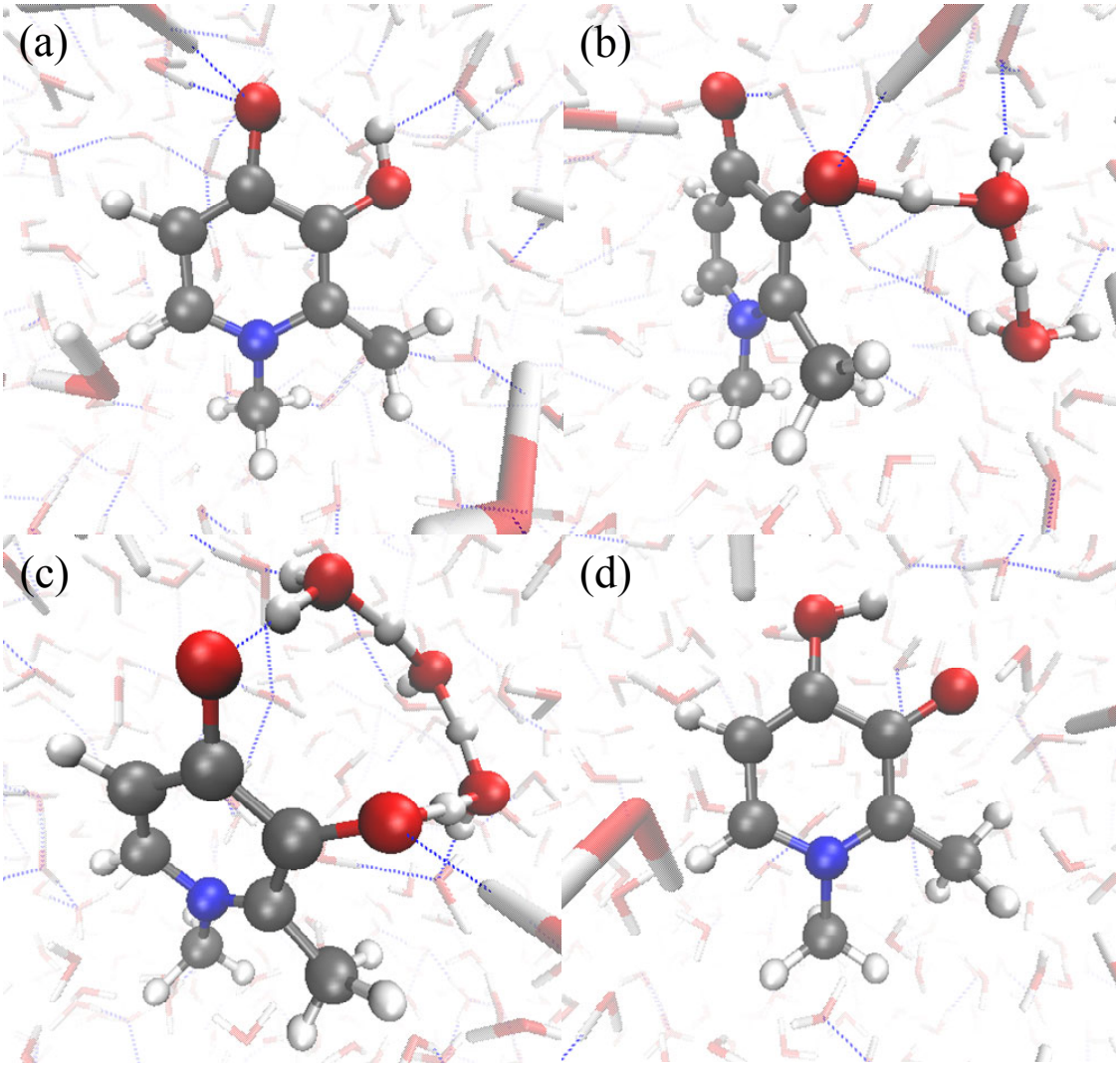


Figure 4-3 Snapshots from the simulation of hydrated CP20 illustrating the deprotonation process induced by the coordination constraint.

(a) n = 0.97: the proton is covalently bonded to the O_3 atom with a hydrogen bond to a nearby H_2O molecule. (b) n = 0.71: the proton is shared between the O_3 atom and the nearby H_2O molecule (c) n = 0.31: a chain of three H_2O molecules forms a bridge mediating a proton transfer chain reaction between the O_3 and O_4 atoms (d) n = 0.01: the O_3 atom is completely deprotonated, while the O_4 atom is now protonated.

Table 4-1 pK_a prediction results using different methods.

	CP60	CP20	CP751	Average Absolute Error
Measured value	8.79[a]	9.76[a]	10.32[b]	N/A
CP-MD	9.00	9.36	10.73	0.34
ACD/pK_a DB 12.0	9.59	9.90	10.03	0.41
Marvin 5.4.0.1	10.79	10.72	10.71	1.12
The direct method using B3LYP/6-31+G(d)/CPCM	14.32	15.90	17.36	6.25
Proton exchange scheme using B3LYP/6-31+G(d)/CPCM	9.07	10.65	12.11	0.99
QM static calculations with a regression model	8.81	9.76	10.18	0.06

Experimental pK_a values taken from [a] Kong (2008); [b] Chen *et al.* (2012a) or Chapter 2.

4.4 Discussion

Although the CP-MD method yields only the second best predictions with respect to average absolute error among all the methods described in this study, it does provide valuable additional insights into the atomistic details of the proton transfer mechanism and the hydration structures and dynamics at all stages of the reaction. Moreover, the CP-MD method is an *ab initio* MD method, meaning that no experimental data are required and this can be superior to the QM static calculations with a regression model for predictions involving novel compounds with few existing data.

The distances of the transferring hydrogen to the donor oxygen (O_3 of HPO) and the acceptor oxygen (O_A of H_2O) *versus* O_3 proton coordination number (n) for all three HPOs were analysed (Figure 4-4). Near thermal equilibrium (n ~ 0.97), the O_3H bond length is observed to be at its equilibrium value of about 0.99 Å for all three HPOs, while the $H\cdots O_A$ hydrogen bond length exhibits significant variation. CP60 is observed to have the shortest average hydrogen bond length of around 1.7 Å, in an agreement with the fact that it possesses the lowest pK_a value. CP751 is observed to have the longest average hydrogen bond length of about 2.1 Å, in agreement with the fact that it possesses the highest pK_a value. CP20 is observed to lie between CP60 and CP751 with respect to the average hydrogen bond length and pK_a value. These differences in hydrogen bond length among the three HPOs may be explained by the differences in the water hydrogen bonding networks surrounding their O_3 atoms – which arise because of their diverse patterns of methyl substitution which lead to differing numbers of water molecules within a radius of about 3 Å from O_3 (Figures 4-5b and 4-6a). Interestingly, for all three HPOs, the crossover point at which the transferring H is midway between O_3 and O_A

coincides at n = 0.72 where the OH distance is around 1.25 Å. After the transition state, the curves for the three HPOs diverge slightly, indicating that CP60 forms the tightest HO_A bond, followed by CP20 and CP751.

Radial distribution functions (RDFs) between the O_3 and water oxygens were calculated to search for a correlation between the pK_a value of HPO and its hydration properties. The O_3–O RDFs for the three HPOs at n = 0.97 are presented in Figure 4-5. The first RDF peak of CP60 is much higher and the separation between the first and second peaks is much more pronounced than for the other two HPOs, indicating a much stronger solute – solvent interaction compatible with its lowest pK_a value. Although the first peak for CP751 is higher than that for CP20, the number of water molecules within a radius of about 4 Å from its O_3 is significantly lower because of the 5-methyl substituent (Figures 4-5b and 4-6a).

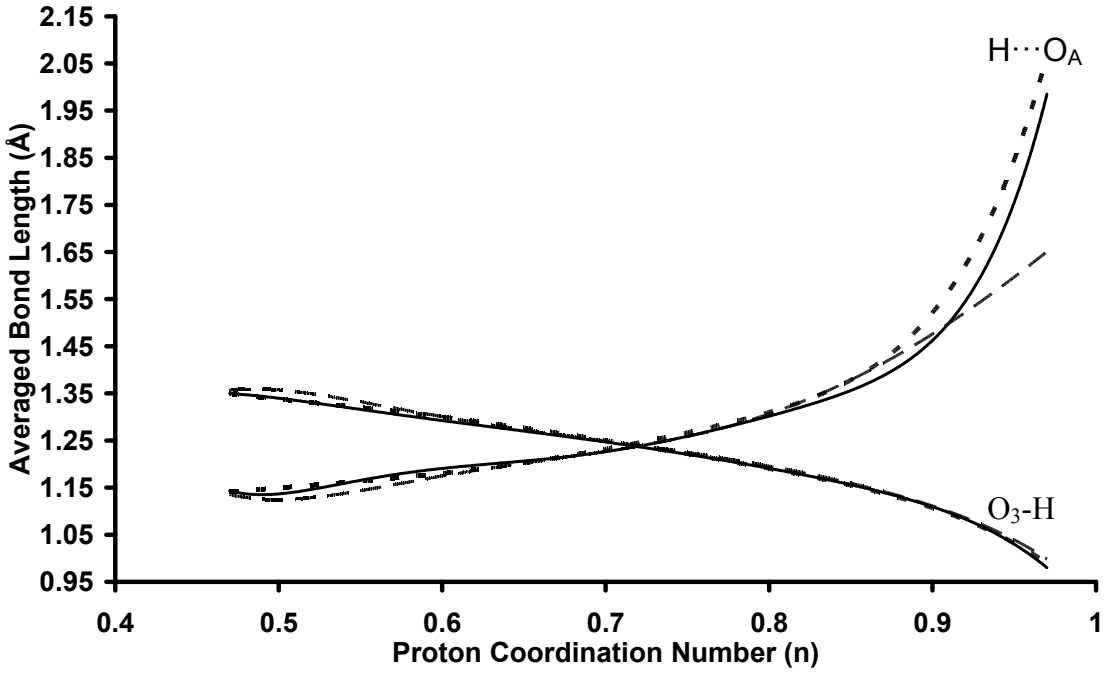


Figure 4-4 Mean OH distances between the transferring H, and the donor (O_3 of HPO) and acceptor (O_A of H_2O) oxygen *versus* O_3 proton coordination number for the three HPOs[a].

[a] CP20 (solid line), CP60 (dashed line), CP751 (short-dashed line).

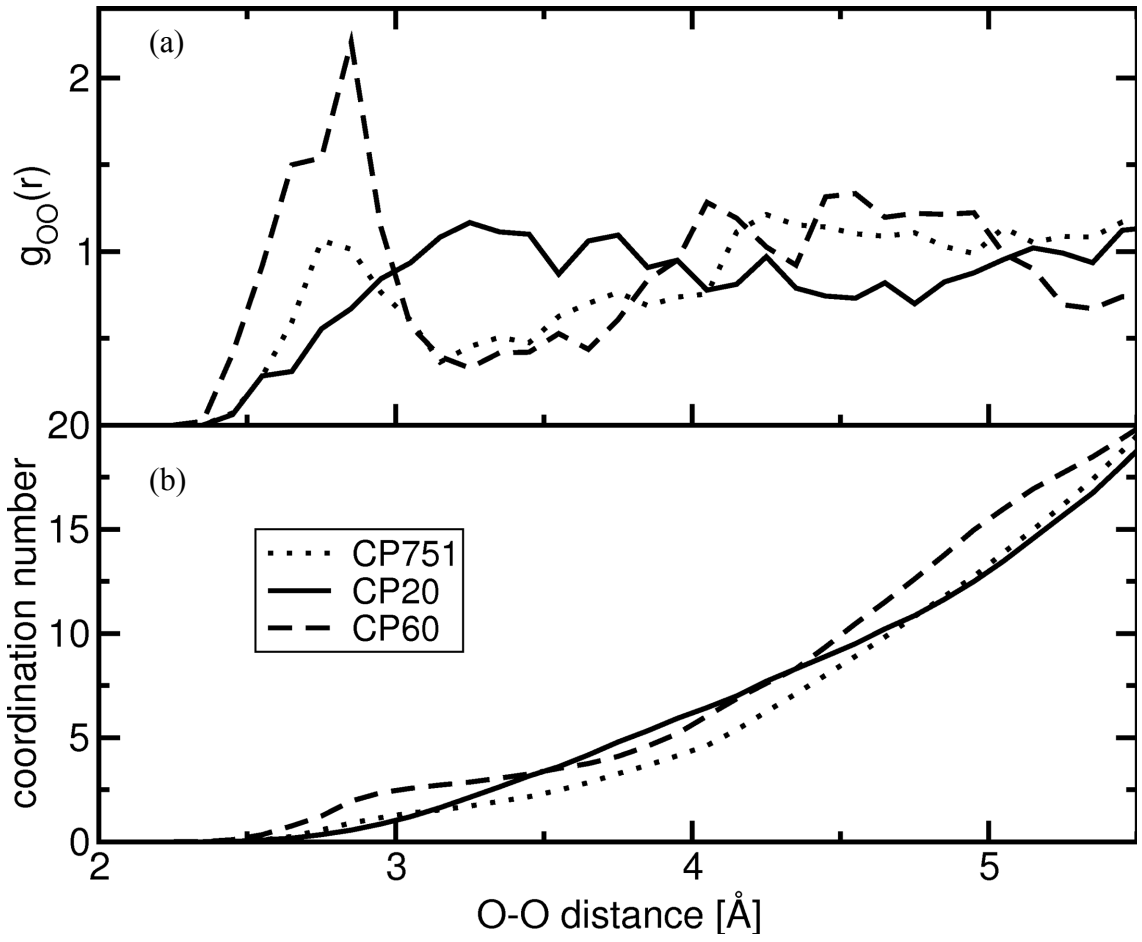


Figure 4-5 O_3–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated HPOs[a] at a mean proton coordination number of 0.97.

[a] CP20 (solid line), CP60 (dashed line), CP751 (short-dashed line).

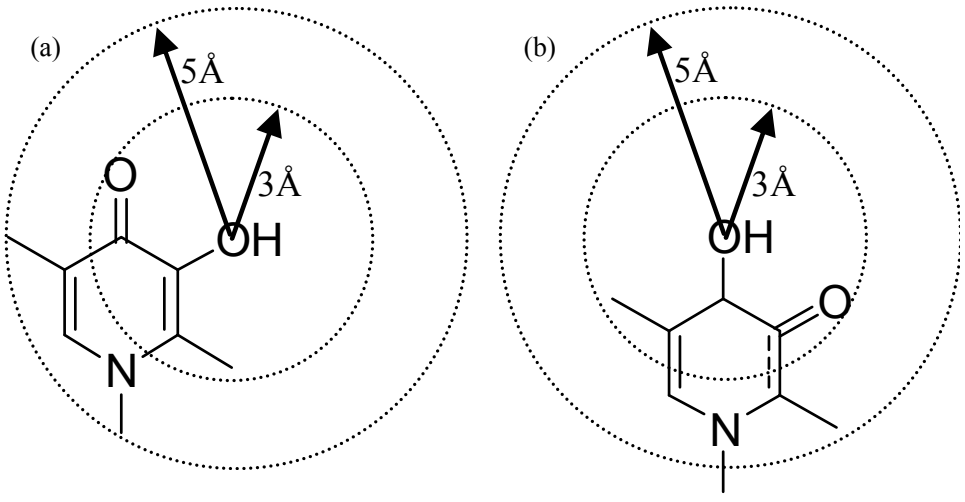


Figure 4-6 Schematic diagram illustrating the 3 Å and 5 Å HPO coordination radii with origin (a) at O_3; (b) at O_4.

In addition to providing atomistic level insights into the hydration and proton transfer properties of the different HPOs, the simulations also offer the possibility to characterise the final product state after proton transfer. Using the methodology applied in this work, it would be possible to determine the pK_a values of the carbonyl oxygen O_4, which is fully protonated at the end of the simulation ($n \rightarrow 0$). However, a quantitative analysis of this was not attempted here. Instead, attempts were made to rationalise the experimentally determined pK_a values of 3.66, 3.38, and 3.28 (Kong, 2008) for CP20, CP60 and CP751, respectively, on the grounds of their hydration structures. The O_4–O RDFs in Figure 4-7a indicate that CP751 forms stronger hydrogen bonds at O_4, the first peak appearing at a shorter distance compared to the other two HPOs, which may explain the slightly lower pK_a value of CP751. The RDF for CP20 exhibits a much more pronounced dip between the first and second peaks which means that the first and second hydration shells are well separated. Thus, the integrated H_2O coordination number around O_4 for CP20 is significantly lower than that for the other two HPOs up to a distance of about 4.5 Å (Figures 4-7b and 4-6b), and this may correlate with the increased pK_a value of CP20. Comparison of Figures 4-5b and 4-7b indicates that the differences in the H_2O coordination number around O_4 are much smaller than those around O_3, explaining the smaller differences in the pK_a values of the O_4H group.

In summary, the CP-MD method provides predicted hydroxyl pK_a values comparable in accuracy with those derived through the QM static method plus experimental fitting described in Chapter 3, and reveals that the proton transfer for all three HPOs takes place along a chain of three H_2O molecules although direct hydrogen bonds are observed to form transiently. The distribution of methyl substitution around the aromatic HPO ring seems to affect the ambient water networking significantly and can partially account for the pK_a value differences.

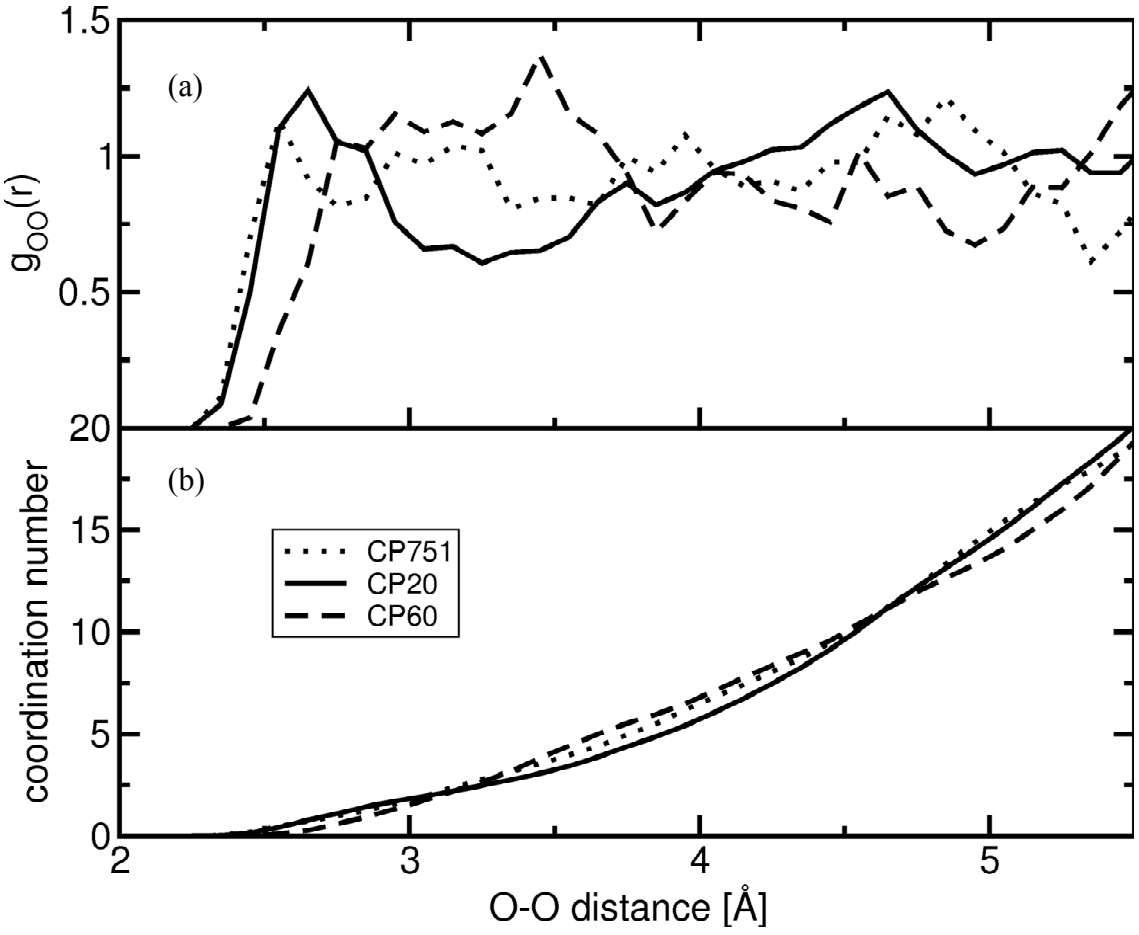


Figure 4-7 O_4–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated HPOs[a] at a mean proton coordination number of 0.01, complete proton transfer from O_3 to O_4.

[a] CP20 (solid line), CP60 (dashed line), CP751 (short-dashed line).

Chapter 5 Prediction of HPO log K_1 Values

5.1 Introduction

log K_1 is another important physicochemical property of iron (Fe^{3+}) chelators, closely related to the coordinating functional group pK_a values, and one of the variables influencing pFe^{3+}. The prediction of this was made using QM static calculations, employing a methodology analogous to that described in Chapter 3. Prior to the prediction, the experimental HPO log K_1 values were verified *via* their correlation with the corresponding sum of the coordinating functional group experimental pK_a values. The prediction methodology reported here is the first to be provided for predicting the absolute log K_1 values of iron-chelating agents in the absence of pK_a values. The main results from this are published in Chen *et al.* (2012b).

5.2 Methods

5.2.1 Prediction of HPO log K_1 values using QM static calculations

In order to simplify calculations and to decrease computing time, only two types of optimised structures, namely deprotonated ligand (L^-) and iron (Fe^{3+}) complex ($[Fe^{3+}L_1]^{2+}$), in solvent models were computed (as analogous to the related studies detailed in Chapter 3 and in Chen *et al.* (2012a)). The calculated free energy differences ($\Delta G^*_{calculated}$) between the two optimised structures of the eleven training set HPOs (Table 5-1) were plotted against their experimentally determined log K_1 values in order to assess the correlation. The correlation coefficient (r) for the regression between predicted and experimental values and their mean and standard deviation of absolute deviations ($|M|$ and $|S|$) were used to investigate the performance of various model chemistries. The different model chemistries at B3LYP/6-31+G(d) level explored here vary very significantly according to the solvent model employed (IEF-PCM, CPCM, SMD, I-PCM or SCI-PCM) (Foresman and Frisch, 1996) and the radii employed in the CPCM models (Bonding, Pauling, UFF, UAHF, or UAKS). For $[Fe^{3+}L_1]^{2+}$, there are two constructed starting structures: one, without explicit water molecules (Figure 5-1a) and the other, with four explicit water molecules octahedrally distributed around the Fe^{3+} (Figure 5-1b). All the optimised geometries (directly optimised in solvent models; force constants calculated at each step; without symmetry constraints; the spin multiplicity = 6 for $[Fe^{3+}L_1]^{2+}$) were calculated using Gaussian 09 (Frisch *et al.*, 2009), and the conformers generated were verified as corresponding to local minima on their potential energy surfaces (with no imaginary frequencies existing in the output files). Prior

to running the Gaussian 09 optimisation, conformational searches at AM1 level were performed using HyperChem Release 8 (http://www.hyper.com/) to obtain the global (or nearly global) minimum energy conformer for all the molecules studied here. (Any further calculations including, for example, Boltzmann averaging over multiple low energy conformers were not performed since the additional computing time required could not be justified.)

The model chemistry which produced the best result was found to be B3LYP/6-31+G(d)/CPCM(Bondi radii, water as solvent) with $[Fe^{3+}L_1]^{2+}$ structures modelled in the absence of explicit water molecules (r^2 = 0.90, $|M|$ = 0.16 and $|S|$ = 0.09; (Figure 5-2)). Adopting the same solvent model, B3LYP*/6-31+G(d) (which has 10%, 15%, 25%, 30%, 35%, or 40% Hartree-Fock exchange compared to 20% for B3LYP), CAM-B3LYP/6-31+G(d) (a modified functional which is suitable for long range calculations), B3LYP/3-21G (a lower basis set), PM6 (a semi-empirical method), were also tested. However, no superior outcomes were detected with respect to either computing accuracy or time.

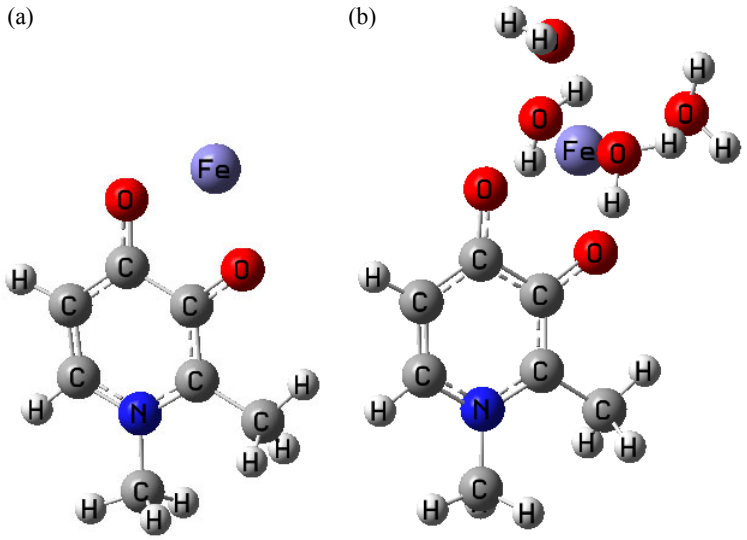


Figure 5-1 $[Fe^{3+}L_1]^{2+}$ Structures: (a) with no explicit water molecules; (b) with four explicit water molecules octahedrally distributed around Fe^{3+}.

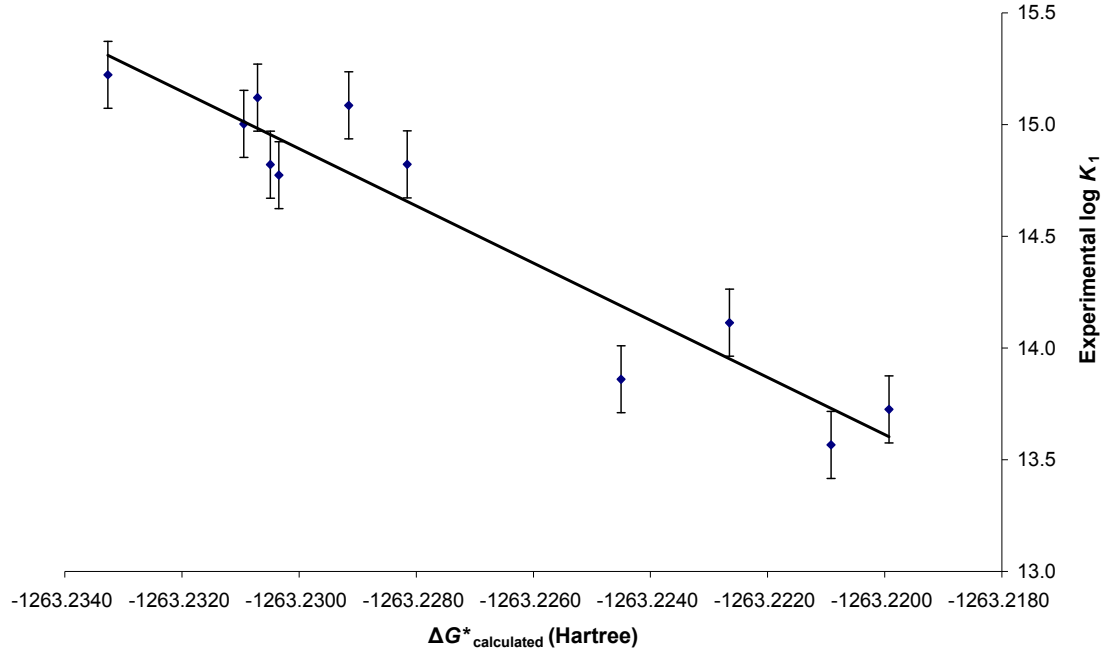
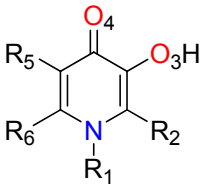


Figure 5-2 Correlation of experimental log K_1 values[a] against the calculated free energy differences ($\Delta G^*_{calculated}$) for the eleven training set HPOs using B3LYP/6-31+G(d)/CPCM(Bondi radii, water as solvent).

[a] The errors related to the experimental log K_1 determination for the typical HPOs are estimated to be of the order of ± 0.15 (see discussion in Chapter 2). 1 Hartree = 627.5095 kcal mol^{-1}

Table 5-1 Structures, predicted (experimental) log K_1 values and calculated free energy differences ($\Delta G^*_{calculated}$) of 3-hydroxypyridin-4-ones (HPOs).



HPOs	R_1	R_2	R_5	R_6
CP20[a]	$-CH_3$	$-CH_3$	−H	−H
CP21[a]	$-C_2H_5$	$-CH_3$	−H	−H
CP23[a]	$-CH(CH_3)_2$	$-CH_3$	−H	−H
CP60[a]	$-CH_3$	−H	−H	−H
CP61[a]	$-C_2H_5$	−H	−H	−H
CP68[a]	$-CH_3$	$-CH_3$	−H	$-CH_3$
CP69[a]	$-CH_3$	−H	−H	$-CH_3$
CP90[a]	−H	−H	−H	−H
CP93[a]	$-CH_3$	$-C_2H_5$	−H	−H
CP94[a]	$-C_2H_5$	$-C_2H_5$	−H	−H
CP99[a]	−H	$-C_2H_5$	−H	−H
CP360	$-CH_3$	$-CH_2OH$	−H	$-CH_3$
CP369	$-C_2H_5$	$-CH_2OH$	−H	$-CH_3$
CP374	$-CH_3$	$-CHC_2H_5OH$	−H	$-CH_3$
CP502	$-CH_3$	$-CONHCH_3$	−H	$-CH_3$
CP509	−H	$-CON(CH_3)_2$	−H	$-CH_3$
CP513	$-CH_3$	$-CONHC_2H_5$	−H	$-CH_3$
CP514	$-C_2H_5$	$-CONHCH_3$	−H	$-CH_3$
YMF5	−H	−H	−F	−H
YMF7	$-CH_3$	−F	−H	−H
YMF8	$-C_2H_5$	−F	−H	−H
YMF14	$-CH_3$	−F	$-CH_3$	−H
YMF15	$-CH_3$	−H	−F	−H
YMF16	$-C_2H_5$	−H	−F	−H
YMF17	$-CH_3$	$-CH_3$	−F	−H
YMF21	−H	−H	−F	$-CH_3$
YMF22	−H	$-CH_3$	−F	−H
YMF24	$-C_3H_7$	−F	−H	−H
YMF25	$-C_3H_7$	−H	−F	−H
YMF26	$-CH(CH_3)_2$	−H	−F	−H
YMF33	$-C_4H_9$	−F	−H	−H
CP102	$-C_2H_4OH$	$-C_2H_5$	−H	−H
CP110	$-C_2H_4COOH$	$-C_2H_5$	−H	−H
CP359	$-C_2H_5$	$-CH_2OH$	−H	−H
CP364	$-CH_3$	$-CH_2OH$	−H	−H
CP365	$-C_2H_5$	$-CHCH_3OH$	−H	−H
CP366	$-C_3H_6OH$	$-CH_2OH$	−H	−H
CP372	$-C_3H_6OH$	$-CHCH_3OH$	−H	−H
CP511	−H	$-CONHCH_3$	−H	$-CH_3$
CP545	−H	$-CONHC_2H_4OH$	−H	−H
CP510	$-CH_3$	−H	−H	$-CONHCH_3$
CP28	−H	$-CH_3$	−H	−H
CP38	$-C_2H_4COOH$	$-CH_3$	−H	−H
CP40	$-C_2H_4OH$	$-CH_3$	−H	−H
CP111	$-C_3H_6COOH$	$-C_2H_5$	−H	−H
CP352	$-C_2H_4OH$	$-CH(C_6H_5)OH$	−H	$-CH_3$
CP362	$-CH_3$	$-CH_2OCH_3$	−H	$-CH_3$
CP363	$-CH_3$	$-CH(CH_3)OCH_3$	−H	$-CH_3$
CP414	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_3$
CP417	$-CH_3$	$-CH_2(NC_5H_{10})$	−H	$-CH_2OH$
CP751	$-CH_3$	$-CH_3$	$-CH_3$	−H
YMF3	−H	−F	−F	−F
YMF4	−H	−F	−H	−F
CP70	−H	−H	−H	$-CH_3$
CP370	$-C_2H_5$	$-CH_2OCH_3$	−H	$-CH_3$
CP375	$-CH_3$	$-CHC_2H_5OCH_3$	−H	$-CH_3$
CP616	−H	$-CH_3$	$-CH_3$	−H

Table 5-1 continued

HPOs	$\Delta G^*_{calculated}$ ($G^*_{calculated}$ for $[Fe^{3+}L_1]^{2+}$, L^-) (Hartree)[g]	log K_1
CP20[a]	-1263.2292 (-1740.0518, -476.8226)	14.79 (15.09)[b]
CP21[a]	-1263.2305 (-1779.3418, -516.1113)	14.98 (14.82)[b]
CP23[a]	-1263.2309 (-1818.6296, -555.3987)	15.04 (15.00)[b]
CP60[a]	-1263.2209 (-1700.7550, -437.5340)	13.64 (13.57)[b]
CP61[a]	-1263.2227 (-1740.0471, -476.8244)	13.88 (14.11)[b]
CP68[a]	-1263.2333 (-1779.3443, -516.1110)	15.37 (15.22)[b]
CP69[a]	-1263.2245 (-1740.0498, -476.8253)	14.14 (13.86)[b]
CP90[a]	-1263.2199 (-1661.4703, -398.2503)	13.50 (13.73)[b]
CP93[a]	-1263.2303 (-1779.3381, -516.1078)	14.96 (14.77)[b]
CP94[a]	-1263.2307 (-1818.6283, -555.3975)	15.01 (15.12)[b]
CP99[a]	-1263.2282 (-1740.0596, -476.8314)	14.65 (14.82)[b]
CP360	-1263.2241 (-1854.5518, -591.3277)	14.09 (14.39)[c]
CP369	-1263.2242 (-1893.8405, -630.6162)	14.10 (14.37)[c]
CP374	-1263.2212 (-1933.1295, -669.9084)	13.67 (14.41)[c]
CP502	-1263.2174 (-1948.0243, -684.8069)	13.14 (13.41)[d]
CP509	-1263.2169 (-1948.0272, -684.8103)	13.08 (12.85)[d]
CP513	-1263.2164 (-1987.3136, -724.0972)	13.00 (12.84)[d]
CP514	-1263.2183 (-1987.3120, -724.0937)	13.28 (12.96)[d]
YMF5	-1263.2103 (-1760.7086, -497.4982)	12.15 (11.57)[e]
YMF7	-1263.2141 (-1799.9948, -536.7807)	12.69 (11.95)[e]
YMF8	-1263.2146 (-1839.2858, -576.0712)	12.75 (11.98)[e]
YMF14	-1263.2178 (-1839.2906, -576.0728)	13.20 (13.14)[e]
YMF15	-1263.2117 (-1799.9937, -536.7820)	12.35 (12.14)[e]
YMF16	-1263.2129 (-1839.2857, -576.0728)	12.51 (12.69)[e]
YMF17	-1263.2196 (-1839.2903, -576.0706)	13.46 (13.61)[e]
YMF21	-1263.2131 (-1800.0077, -536.7946)	12.54 (12.50)[e]
YMF22	-1263.2170 (-1800.0099, -536.7929)	13.09 (13.21)[e]
YMF24	-1263.2144 (-1878.5745, -615.3601)	12.73 (12.20)[e]
YMF25	-1263.2124 (-1878.5741, -615.3617)	12.44 (11.88)[e]
YMF26	-1263.2133 (-1878.5751, -615.3618)	12.57 (11.76)[e]
YMF33	-1263.2147 (-1917.8633, -654.6486)	12.77 (12.37)[e]
CP102	-1263.2278 (-1893.8415, -630.6136)	14.61 (14.69)[f]
CP110	-1263.2271 (-2007.2015, -743.9744)	14.50 (14.78)[f]
CP359	-1263.2226 (-1854.5508, -591.3281)	13.88 (14.26)[f]
CP364	-1263.2206 (-1815.2591, -552.0386)	13.59 (14.28)[f]
CP365	-1263.2197 (-1893.8404, -630.6207)	13.47 (14.15)[f]
CP366	-1263.2203 (-1969.0530, -705.8327)	13.55 (13.82)[f]
CP372	-1263.2195 (-2008.3458, -745.1263)	13.43 (14.29)[f]
CP511	-1263.2063 (-1908.7534, -645.5471)	11.59 (12.03)[f]
CP545	-1263.2005 (-1983.9593, -720.7588)	10.77 (11.32)[f]
CP510	-1263.2172 (-1908.7311, -645.5139)	13.12 (12.20)[f]
CP28	-1263.2268 (-1700.7715, -437.5446)	14.46 (14.49)[f]
CP38	-1263.2265 (-1967.9156, -704.6891)	14.42 (14.62)[f]
CP40	-1263.2280 (-1854.5560, -591.3280)	14.63 (14.61)[f]
CP111	-1263.2276 (-2046.4899, -783.2623)	14.57 (14.58)[f]
CP352	-1263.2162 (-2200.0370, -936.8208)	12.97 (13.02)[f]
CP362	-1263.2249 (-1893.8275, -630.6026)	14.19 (14.45)[f]
CP363	-1263.2280 (-1933.1183, -669.8903)	14.64 (15.00)[f]
CP414	-1263.2056 (-2030.3502, -767.1446)	11.49 (11.59)[f]
CP417	-1263.2013 (-2105.5593, -842.3581)	10.88 (10.98)[f]
CP751	-1263.2338 (-1779.3475, -516.1137)	15.44 (15.51)[f]
YMF3	-1263.2358 (-1958.7859, -695.5501)	15.72 (15.06)[f]
YMF4	-1263.2455 (-1859.5496, -596.3041)	17.08 (16.52)[f]
CP70	-1263.2226 (-1700.7691, -437.5465)	13.87 (14.09)[f]
CP370	-1263.2257 (-1933.1174, -669.8917)	14.31 (14.29)[f]
CP375	-1263.2279 (-1972.4049, -709.1770)	14.62 (14.50)[f]
CP616	-1263.2309 (-1740.0669, -476.8360)	15.04 (15.17)[f]

[a] The first eleven HPOs were used to test different model chemistries; experimental log K_1 values taken from [b] Kong (2008); [c] Liu *et al.* (1999); [d] Piyamongkol *et al.* (2010); [e] Ma *et al.* (2012); [f] this study obtained from Chapter 2; [g] $\Delta G^*_{calculated}$: the calculated free energy differences between $[Fe^{3+}L_1]^{2+}$ and L^- using B3LYP/6-31+G(d)/CPCM(Bondi radii, water as solvent) ($G^*_{calculated}$: the calculated free energy); 1 Hartree = 627.5095 kcal mol^{-1}.

5.3 Results

5.3.1 Validation of experimental log K_1 values

In order to obtain a large group of HPOs with reliable experimentally determined log K_1 values, a plot of experimental log K_1 value *versus* the corresponding sum of experimental pK_a values (hydroxyl pK_a values plus carbonyl pK_a values) was utilised. From the existing literature (Kong, 2008; Liu *et al.*, 1999; Ma *et al.*, 2012; Piyamongkol *et al.*, 2010; Rai *et al.*, 1998), forty-one HPOs were selected for initial analysis. Thirty one of these HPOs were found to give the expected linear correlation (Hider and Hall, 1991) between their log K_1 values and the corresponding sum of their pK_a values, r^2 = 0.95 (Figure 5-3). However, nine HPOs were found to have significantly higher log K_1 values than predicted from the linear relationship, while one HPO was found to possess a significantly lower log K_1 value (Figure 5-3). After determining the affinity constants using a 10 : 1 molar ratio of L : Fe^{3+} for the ten HPO outliers (Table 5-1 or Table 5-2 including former values) and sixteen previously uncharacterised HPOs (Table 5-1), four of which were specifically synthesised for this study by Ma (Chen *et al.*, 2012b), fifty-seven HPOs in total were found to give the expected linear relationship between their experimental log K_1 values and the corresponding sum of experimental pK_a values, r^2 = 0.95 (Figure 5-4).

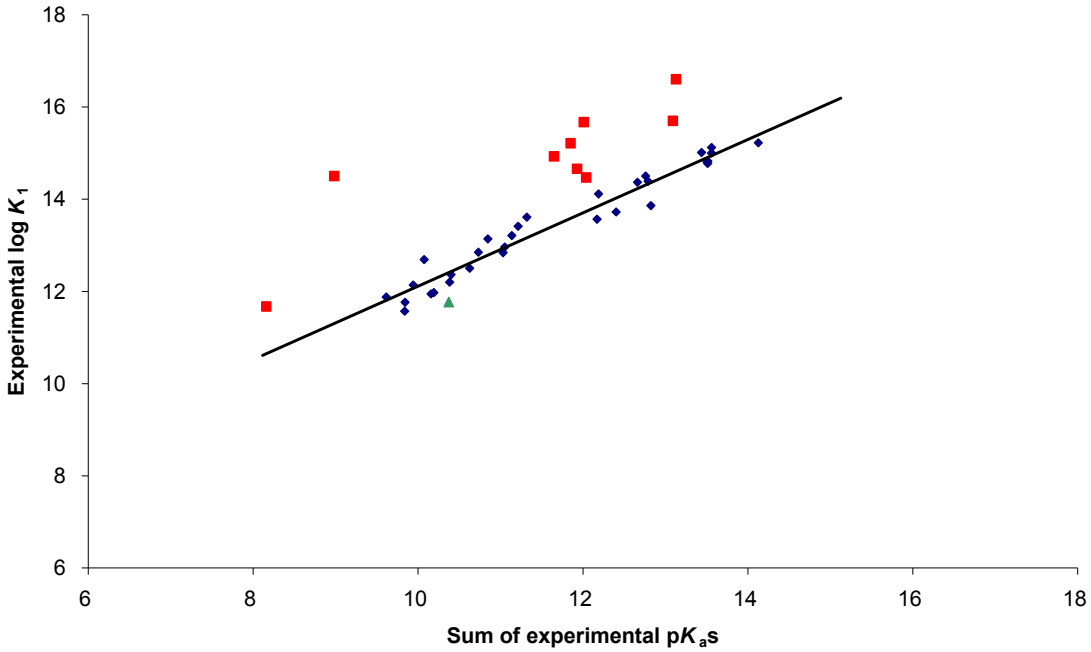


Figure 5-3 Correlation of experimental log K_1 values against the sum of experimental pK_a values for forty-one HPOs[a], extracted from the existing literature[b].

■ CP102, CP110, CP359, CP364, CP365, CP366, CP372, CP511, CP545; ▲ CP510;
[a] the structures of the forty-one HPOs are presented in the first forty-one rows of Table 5-1. [b](Kong, 2008; Liu *et al.*, 1999; Ma *et al.*, 2012; Piyamongkol *et al.*, 2010; Rai *et al.*, 1998) using a 3.5 : 1 molar ratio of L : Fe^{3+}.

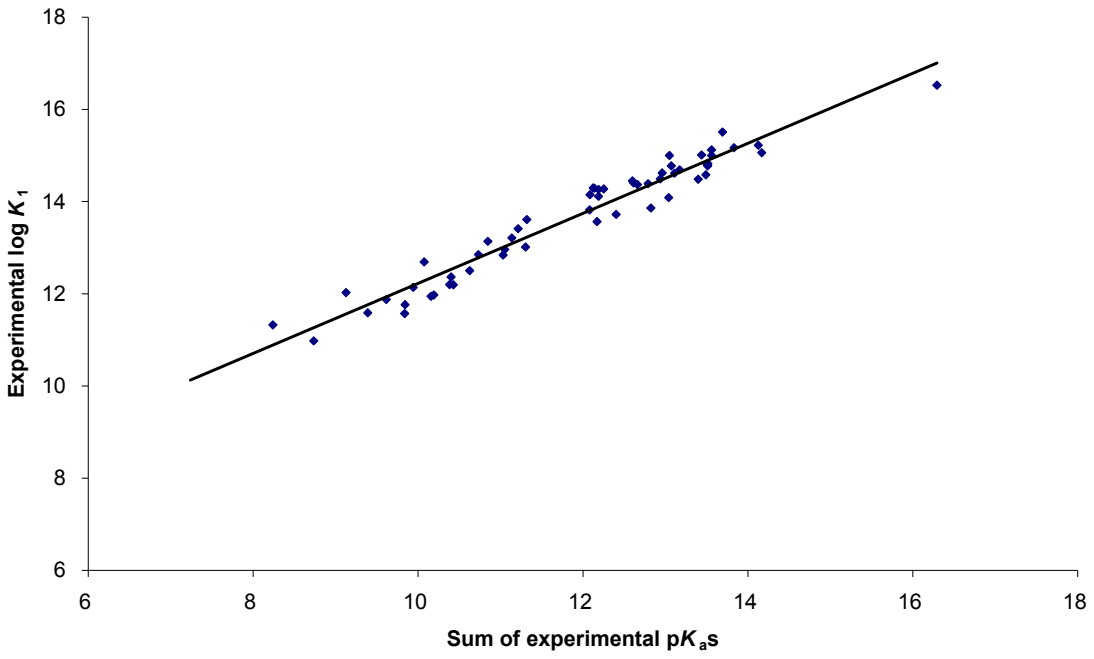


Figure 5-4 Correlation of experimental log K_1 values against the sum of experimental pK_a values for fifty-seven HPOs[a].

[a] The structures of the fifty-seven HPOs are presented in Table 5-1. Ten outliers in Figure 5-3 were re-determined using a 10 : 1 molar ratio of L : Fe^{3+}.

Table 5-2 Re-determined log K_1 values of the ten HPO outliers[a] in Figure 5-3.

	Log K_1
CP102	14.69 (15.70)[b]
CP110	14.78 (16.60)[b]
CP359	14.26 (14.66)[c]
CP364	14.28 (14.47)[c]
CP365	14.15 (15.21)[c]
CP366	13.82 (15.67)[c]
CP372	14.29 (14.93)[c]
CP511	12.03 (14.50)[d]
CP545	11.32 (11.67)[d]
CP510	12.20 (11.77)[d]

[a] The structures of these HPOs are given in Table 5-1; experimental log K_1 values taken from [b] Rai *et al.* (1998); [c] Liu *et al.* (1999); [d] Piyamongkol *et al.* (2010).

5.3.2 Developed log K_1 prediction model

As discussed above, B3LYP/6-31+G(d)/CPCM(Bondi radii, water as solvent), with $[Fe^{3+}L_1]^{2+}$ structures being in the absence of explicit water molecules, was found to produce a relatively reliable prediction of log K_1 values for an initial series of eleven HPOs. The log K_1 values for an additional forty-two HPOs were calculated using the same model chemistry to develop the final log K_1 prediction model — Predicted log K_1 = -140.18 $\Delta G^*_{calculated}$ - 177063.32 (r^2 = 0.90, $|M|$ = 0.32 and $|S|$ = 0.25, Figure 5-5; values in Table 5-1). In an attempt to further examine the predictive power of this model, four novel HPOs, CP70, CP370, CP375 and CP616, were

synthesised by Ma (Chen *et al.*, 2012b). The predicted log K_1 values for all four compounds were observed to be close to their experimental values, indicating reliable predictions (Figure 5-5). The absolute deviations between the predicted and experimental values are 0.22 for CP70 (13.87 *vs.* 14.09), 0.02 for CP370 (14.31 *vs.* 14.29), 0.12 for CP375 (14.62 *vs.* 14.50), and 0.13 for CP616 (15.04 *vs.* 15.17).

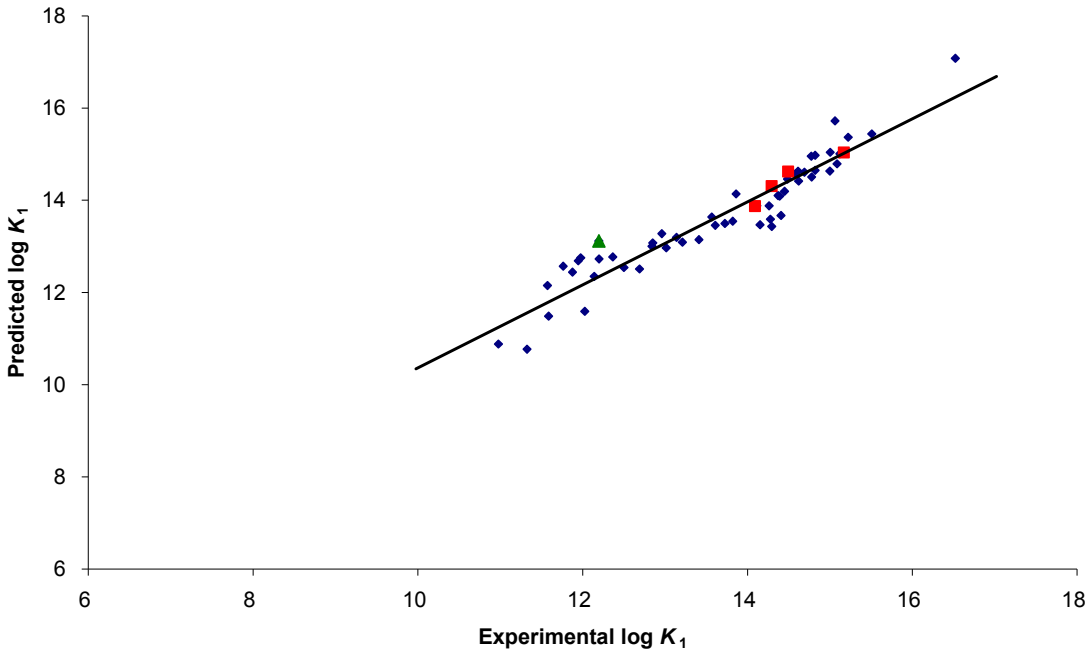


Figure 5-5 Developed log K_1 prediction model.

■ CP70, CP370, CP375 and CP616; ▲CP510 with the largest absolute deviation (0.92); Omission of the CP510 data point changes only the third decimal place in r^2 and results in an insignificant change of the fitted line, and thus an insignificant change of the predicted log K_1 values.

5.4 Discussion

The methodology developed for prediction of HPO log K_1 values is impressive (with the observed and predicted values strongly correlated, with r^2 of 0.90; Figure 5-5). In comparison, the errors associated with the experimental determination of log K_1 for the typical HPOs are estimated to be of the order of ± 0.15 (see the discussion in Chapter 2). However, for some HPOs which are associated with experimentally difficult log K_1 determinations, the error can exceed 1.0 (Table 5-2). The plot of experimental log K_1 *versus* the corresponding sum of experimental pK_a values (Figure 5-4) was utilised to evaluate this experimental error. The utilisation of the sum of experimental pK_a values rather than the single carbonyl or hydroxyl experimental pK_a value was adopted because both of the pK_a values are associated with functional groups which coordinate Fe^{3+}.

The ten HPO outliers in Figure 5-3 can be categorised into four main subclasses, namely 1-hydrogen-(2 or 6)-amido-, 2-hydroxymethyl-, 1-hydroxy(ethyl or propyl)- and 1-carboxyethyl-HPOs. All these HPOs possess hydrogen bond donor and/or acceptor groups for the aromatic ring substituents and this may result in a disturbance of the first layer of solvation molecules surrounding each iron(Fe^{3+}) complex. This, in turn, will influence the magnitudes of their stability constants; for instance, with CP511, which has strong intramolecular hydrogen bonding (Piyamongkol *et al.*, 2010), there is an appreciable decrease in log K_1 values (Figure 5-6) when increased molar ratios of L : Fe^{3+} are employed in the titration. This effect is probably associated with the partial formation of a μ-oxo bridge species (Fe^{3+}−O−$Fe^{3+}L^{-}_{1}$) at relatively high Fe^{3+} molar concentrations, thus disturbing the UV/Vis spectra. For this reason, the 10 : 1 molar ratio of L : Fe^{3+} was adopted to minimise the formation of μ-oxo bridges during the experimental determination of log K_1 values in this study.

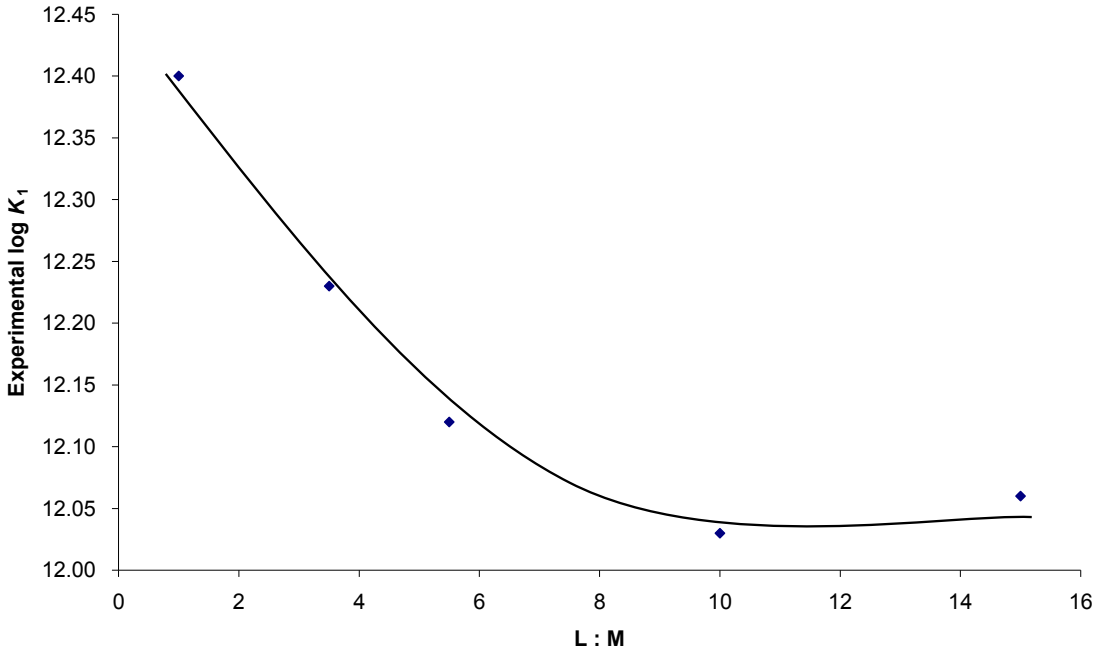


Figure 5-6 Experimental log K_1 *versus* molar ratio of L : M for CP511.

A knowledge of pK_a values can be neglected using the developed prediction methodology for HPO log K_1 values and this is particularly useful to predict HPOs possessing at least three pK_a values. For the experimental log K_1 determination of HPOs with more than two pK_a values, special attention is required to identify which two pK_a values correspond to the iron-coordinating oxygen atoms. If the two pK_a values are incorrectly assigned, there will be an appreciable error in the associated experimental log K_1 value.

In summary, the methodology developed in this study is the first QM approach to predict the absolute log K_1 value of iron-chelating agents in the absence of pK_a values. The plot of experimental log K_1 values *versus* the corresponding sum of experimental pK_a values (of coordinating atoms) has proved to be extremely useful in order to investigate the accuracy of experimentally determined log K_1 values. The stability constants for HPOs with substituents possessing hydrogen bonding donors and/or acceptors should be determined using a 10 : 1 molar ratio of L : Fe^{3+}.

Chapter 6 Prediction of HPO Hydration Property Using Classical Molecular Dynamic Simulations

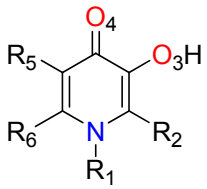
6.1 Introduction

Given, as noted earlier, that HPOs represent putative therapeutic agents for the treatment of neurodegenerative diseases, it is of particular interest to determine their likely blood-brain barrier permeability. In the work reported here, the aim was to develop a suitable *in silico* method to allow modelling of the HPO blood-brain barrier permeation and correlation of the predicted values with the experimental permeability values (Habgood *et al.*, 1999). Through lack of research time, however, it only proved possible to develop the necessary force fields for use in these simulations and then to test these in simulation of the HPO behaviour in aqueous solution. In the future, the developed force fields could be used in simulations including membranes to calculate theoretical blood-brain barrier permeability values for the modelled HPOs.

6.2 Methods

6.2.1 Extension of CHARMM General Force Field for HPOs

The extension of the CHARMM (http://www.charmm.org/) General Force Field (CGenFF) specific for three HPOs (CP20, CP25 and CP94) was made following the procedures given by Vanommeslaeghe *et al.* (2010). Initial estimates of the specific parameters for HPOs with penalty scores indicating their reliability, compared with the structures in CGenFF 2b7 version database, were automatically generated using the utilities provided at https://www.paramchem.org/index.php. Unfortunately, the HPO structures proved significantly different from those in the database (as judged from their penalty scores) and the estimated parameters required optimisation, following the procedure shown in Figure 6-1. In the work reported here, no novel atom types were incorporated and thus, the optimisation of Lennard-Jones parameters, accounting for van der Waals interactions, was not required. The optimisation of bond, angle and dihedral parameters was carried out for all geometries except those covered in the existing CHARMM database. The quantum mechanical reference data were generated using Gaussian 03 (Frisch *et al.*, 2004).



CP20, $R_1 = R_2 = CH_3$ and $R_5 = R_6 = H$
CP25, $R_1 = C_6H_{13}$, $R_2 = CH_3$ and $R_5 = R_6 = H$
CP94, $R_1 = R_2 = C_2H_5$ and $R_5 = R_6 = H$

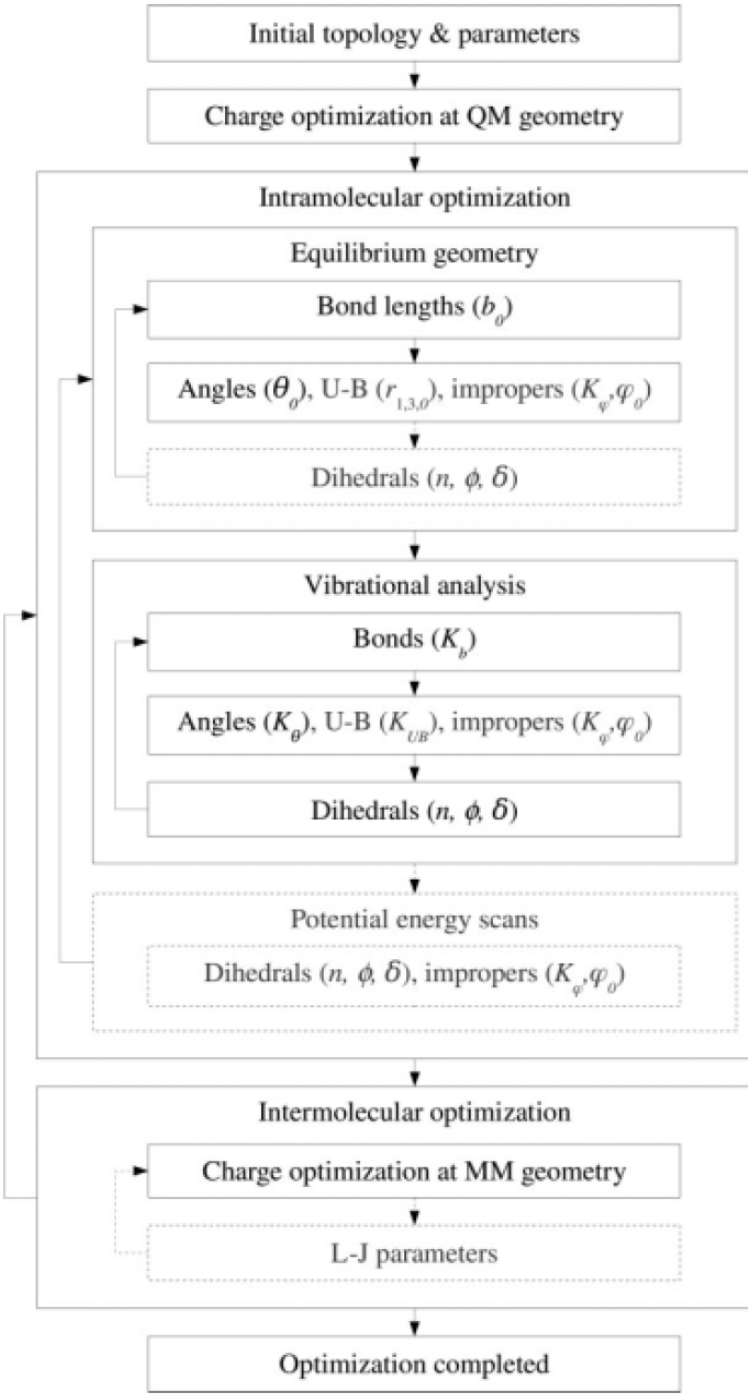


Figure 6-1 Parameterisation procedure (Vanommeslaeghe *et al.*, 2010).

6.2.2 Simulations of HPOs with explicit water molecules using LAMMPS

Once the three novel HPO CHARMM force fields were optimised, they were tested in MD simulations in which each HPO was surrounded by 2,700 H_2O molecules in a periodically repeating cubic unit cell of size 46 Å x 46 Å x 46 Å using LAMMPS (http://lammps.sandia.gov/). The MD simulations were run for three successive 1 ns periods, using NPT conditions (with pressure = 1 atm, temperature = 300 K). The HPO hydration properties were analysed using radial distribution functions (RDFs).

6.3 Results

6.3.1 CHARMM force fields specific for three HPOs

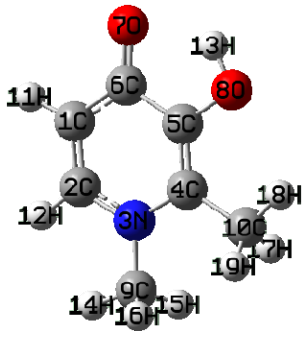
The CHARMM force field parameters developed for three HPOs and their validation results are presented in Tables 6-1 – 6-18 and Figures 6-1 – 6-5. The equilibrium bond lengths, angles, dihedrals and improper dihedrals calculated by QM and CHARMM molecular mechanics (MM) are presented in Tables 6-1 (CP20), 6-7 (CP25) and 6-13 (CP94). The differences of the bond lengths between QM and MM for all three HPOs are within ± 0.03 Å, while those of angles are generally within ± 3^o, both of these internal coordinate sets thus reaching the acceptance criteria specified by Vanommeslaeghe *et al.* (2010). The HPO-water interaction models for all three HPOs are illustrated in Figure 6-2 (interacting atoms: O_3, O_4, H_5 and H_6) and the optimised interaction energies and distances as a result of charge adjustment (Tables 6-5 for CP20, 6-11 for CP25 and 6-17 for CP94) are presented in Tables 6-2 (CP20), 6-8 (CP25) and 6-14 (CP94). The differences of interaction energies between QM and MM for all three HPOs are within ± 0.2 kcal mol^{-1} and the interaction distances of MM are generally 0.2 Å shorter than QM, both the interaction energies and distances thus reaching the acceptance criteria specified by Vanommeslaeghe *et al.* (2010). The dipole moments calculated by QM and MM for the three HPOs are presented in Tables 6-3 (CP20), 6-9 (CP25) and 6-15 (CP94). The dipole moments calculated by MM are generally close to those calculated by QM in this study although it is expected that the MM dipole moment should overestimate the magnitude of the QM one by 20 to 50% for polar, neutral molecules (Vanommeslaeghe *et al.*, 2010). The vibrational spectra calculated by QM and MM as a result of force constants adjustment (Tables 6-6 for CP20, 6-12 for CP25 and 6-18 for CP94) for the three HPOs are presented in Tables 6-4 (CP20), 6-10 (CP25) and 6-16 (CP94). The mean of absolute relative deviations between QM and MM vibrational frequencies for all three HPOs is within 5%, indicating a good agreement between

QM and MM calculations. The potential energy surfaces (PESs) of dihedrals and improper dihedrals calculated by QM and MM for the three HPOs are presented in Figures 6-3 (CP20), 6-4 (CP25) and 6-5 (CP94). The shape of PESs calculated by MM is generally in a good agreement with that calculated through QM. Some of MM PESs have large deviations (greater than 1 kcal mol^{-1}) away from QM ones but the positions of maxima and minima are close to each other.

Table 6-1 CHARMM molecular mechanics (MM) equilibrium geometry of CP20, compared to QM MP2/6-31G(d) level.

Coordinate	MP2	MM	Difference	Coordinate	MP2	MM	Difference
Bond lengths(Å)				Dihedrals(°)			
C1-C2	1.372	1.372	0.001	C6-C1-C2-N3	-0.5	0.0	0.5
C1-C6	1.430	1.429	-0.001	C6-C1-C2-H12	179.8	180.0	0.2
C1-H11	1.085	1.090	0.005	H11-C1-C2-N3	179.8	-180.0	0.2
C2-N3	1.363	1.353	-0.010	H11-C1-C2-H12	0.1	0.0	-0.1
C2-H12	1.086	1.097	0.012	C2-C1-C6-C5	0.5	0.0	-0.5
N3-C4	1.389	1.378	-0.010	C2-C1-C6-O7	-179.8	-180.0	-0.2
N3-C9	1.462	1.469	0.007	H11-C1-C6-C5	-179.7	180.0	-0.3
C4-C5	1.374	1.382	0.008	H11-C1-C6-O7	0.0	0.0	0.0
C4-C10	1.498	1.505	0.007	C1-C2-N3-C4	0.5	0.0	-0.5
C5-C6	1.454	1.463	0.010	C1-C2-N3-C9	178.6	-180.0	1.4
C5-O8	1.356	1.359	0.003	H12-C2-N3-C4	-179.8	180.0	-0.2
C6-O7	1.260	1.235	-0.025	H12-C2-N3-C9	-1.7	0.0	1.7
O8-H13	0.994	0.972	-0.022	C2-N3-C4-C5	-0.7	0.0	0.7
C9-H14	1.090	1.113	0.023	C2-N3-C4-C10	-179.8	180.0	-0.2
C9-H15	1.092	1.114	0.022	C9-N3-C4-C5	-178.7	-180.0	-1.3
C9-H16	1.093	1.114	0.021	C9-N3-C4-C10	2.2	0.0	-2.2
C10-H17	1.095	1.112	0.016	C2-N3-C9-H14	5.5	0.0	-5.5
C10-H18	1.091	1.111	0.020	C2-N3-C9-H15	124.7	119.9	-4.8
C10-H19	1.095	1.112	0.017	C2-N3-C9-H16	-114.0	-119.9	-5.9
Angles(°)				C4-N3-C9-H14	-176.4	180.0	-3.6
C2-C1-C6	120.5	121.3	0.8	C4-N3-C9-H15	-57.2	-60.1	-2.9
C2-C1-H11	119.8	123.7	3.9	C4-N3-C9-H16	64.1	60.1	-4.0
C6-C1-H11	119.6	114.9	-4.7	N3-C4-C5-C6	0.8	0.0	-0.8
C1-C2-N3	122.7	122.8	0.2	N3-C4-C5-O8	-179.6	180.0	-0.4
C1-C2-H12	121.8	121.8	-0.1	C10-C4-C5-C6	179.9	-180.0	0.1
N3-C2-H12	115.5	115.4	-0.1	C10-C4-C5-O8	-0.5	0.0	0.5
C2-N3-C4	120.7	119.0	-1.7	N3-C4-C10-H17	63.8	60.3	-3.5
C2-N3-C9	119.5	119.1	-0.5	N3-C4-C10-H18	-177.2	180.0	-2.8
C4-N3-C9	119.7	121.9	2.2	N3-C4-C10-H19	-58.0	-60.3	-2.3
N3-C4-C5	117.9	121.5	3.5	C5-C4-C10-H17	-115.3	-119.7	-4.4
N3-C4-C10	119.8	119.1	-0.7	C5-C4-C10-H18	3.7	0.0	-3.7
C5-C4-C10	122.3	119.4	-2.8	C5-C4-C10-H19	122.9	119.7	-3.2
C4-C5-C6	123.9	120.8	-3.1	C4-C5-C6-C1	-0.7	0.0	0.7
C4-C5-O8	121.6	124.2	2.6	C4-C5-C6-O7	179.6	-180.0	0.4
C6-C5-O8	114.5	115.0	0.5	O8-C5-C6-C1	179.7	-180.0	0.3
C1-C6-C5	114.3	114.6	0.3	O8-C5-C6-O7	0.0	0.0	0.0
C1-C6-O7	127.7	121.6	-6.1	C4-C5-O8-H13	-179.5	180.0	-0.5
C5-C6-O7	118.0	123.9	5.8	C6-C5-O8-H13	0.1	0.0	-0.1
C5-O8-H13	101.4	101.5	0.1	Improper dihedrals(°)			
N3-C9-H14	108.4	112.5	4.1	C6-C1-C5-O7	-0.2	0.0	0.2
N3-C9-H15	110.5	111.1	0.5				
N3-C9-H16	110.8	111.1	0.2				
H14-C9-H15	108.8	107.0	-1.8				
H14-C9-H16	108.9	107.0	-1.9				
H15-C9-H16	109.2	108.0	-1.3				
C4-C10-H17	112.1	110.8	-1.3				
C4-C10-H18	108.1	112.6	4.5				
C4-C10-H19	112.1	110.8	-1.3				
H17-C10-H18	108.1	106.9	-1.2				
H17-C10-H19	108.1	108.6	0.5				
H18-C10-H19	108.2	106.9	-1.3				

The coordinates highlighted in blue are specific for CP20 and were optimised in this study.



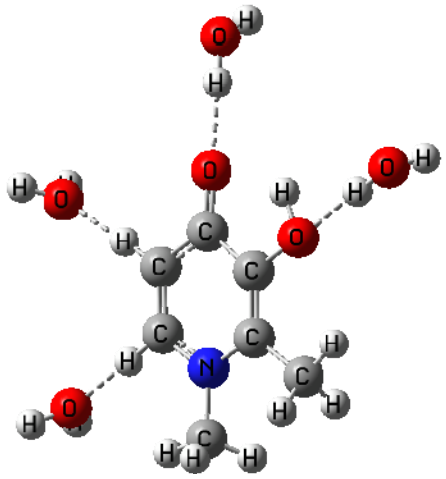


Figure 6-2 Interaction orientations of CP20 with water molecules that were used for charge optimisation[a].

[a] Only a single water molecule is interacting with CP20 during each calculation; all water molecules are shown simultaneously only for convenience. The CP25-water and CP94-water models were set up by analogy with this CP20-water model.

Table 6-2 Interaction energies (kcal mol^{-1}) and distances (Å) of CP20-water complexes in different geometries.

interaction geometry	ΔE(HF)[a]	ΔE(MM)	$\Delta\Delta E$	r(HF)[b]	r(MM)	Δr
O_4···HOH	-7.34	-7.27	0.07	1.99	1.74	-0.25
O_3···HOH	-2.18	-2.20	-0.02	2.23	2.04	-0.19
H_5···OHH	-2.00	-2.04	-0.04	2.53	2.39	-0.14
H_6···OHH	-5.46	-5.52	-0.06	2.35	2.30	-0.05
$\|M\|$			0.05			0.16

[a] HF/6-31G(d) interaction energies are scaled by a factor 1.16 (Vanommeslaeghe *et al.*, 2010); [b] HF interaction distances are not scaled; however, bulk phase hydrogen bonds should be roughly 0.2 shorter than vacuum (Vanommeslaeghe *et al.*, 2010). $|M|$: mean of absolute deviations.

Table 6-3 Dipole moment (Debye) of CP20 calculated at HF/6-31G(d)[a], MP2/6-31G(d) and MM.

μ	HF/6-31G(d)	MP2/6-31G(d)	MM
X	-6.5759	-7.3776	-5.7950
Y	-1.0023	-1.3748	-2.2027
Z	-0.0104	-0.0137	-0.0001
total	6.6519	7.5046	6.1995

[a] The HF dipole moment is calculated on the MP2/6-31G(d) optimised geometry.

Table 6-4 CP20 vibrational spectra calculated at scaled MP2/6-31G(d) and MM.

MP2/6-31G(d)[a]	MM	ΔFrequency%[b]
46.5	41.4	-11.0%
109.9	118.6	7.9%
119.3	131.1	9.9%
174.0	164.4	-5.5%
241.7	287.3	18.9%
302.6	313.3	3.5%
304.0	343.4	13.0%
322.5	362.0	12.2%
348.4	366.6	5.2%
431.6	439.8	1.9%
446.5	464.5	4.0%
474.2	474.1	0.0%
477.4	528.1	10.6%
559.6	558.9	-0.1%
567.7	606.0	6.7%
651.2	647.1	-0.6%
660.2	661.1	0.1%
737.3	716.2	-2.9%
741.6	757.0	2.1%
811.4	820.6	1.1%
874.9	928.3	6.1%
1002.9	965.5	-3.7%
1020.7	1002.4	-1.8%
1037.4	1003.9	-3.2%
1100.8	1010.7	-8.2%
1103.1	1065.4	-3.4%
1147.1	1115.7	-2.7%
1218.3	1200.2	-1.5%
1254.8	1263.4	0.7%
1281.7	1300.3	1.5%
1358.9	1363.5	0.3%
1382.9	1403.7	1.5%
1387.2	1407.4	1.5%
1429.4	1413.7	-1.1%
1448.7	1427.8	-1.4%
1456.0	1437.1	-1.3%
1464.4	1479.1	1.0%
1472.7	1498.3	1.7%
1497.6	1543.6	3.1%
1543.8	1576.1	2.1%
1615.0	1633.0	1.1%
1629.6	1768.2	8.5%
2924.8	2904.9	-0.7%
2943.1	2907.8	-1.2%
2993.7	2957.2	-1.2%
3025.8	2957.7	-2.3%
3042.8	2958.9	-2.8%
3052.5	2959.4	-3.0%
3065.8	2997.5	-2.2%
3084.9	3001.6	-2.7%
3241.9	3675.5	13.4%
	$\|M\|$	4.0%

[a] scaled by a factor of 0.943 (Vanommeslaeghe *et al.*, 2010); [b] ΔFrequency% = (MM_frequency - MP2/6-31G(d)_frequency) (MP2/6-31G(d)_frequency)$^{-1}$ x100%; $|M|$: mean of absolute ΔFrequency%.

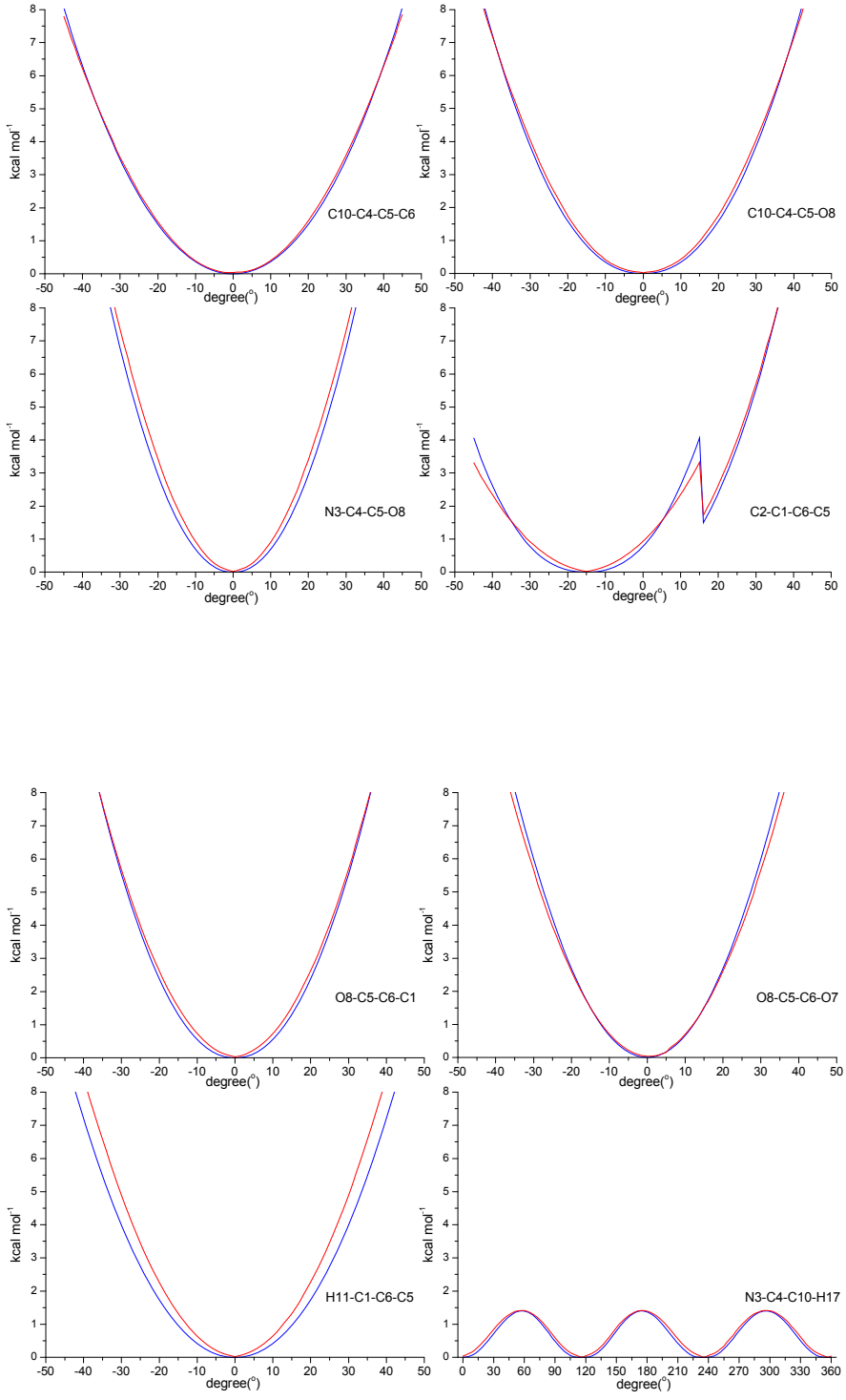


Figure 6-3 Potential energy surfaces of CP20 calculated at MP2/6-31G(d) (blue) and MM (red) for the optimised dihedrals and improper dihedrals.

Y-axis is only shown up to 8 kcal mol^{-1} because this is the only important region for MD simulations at 300 K.

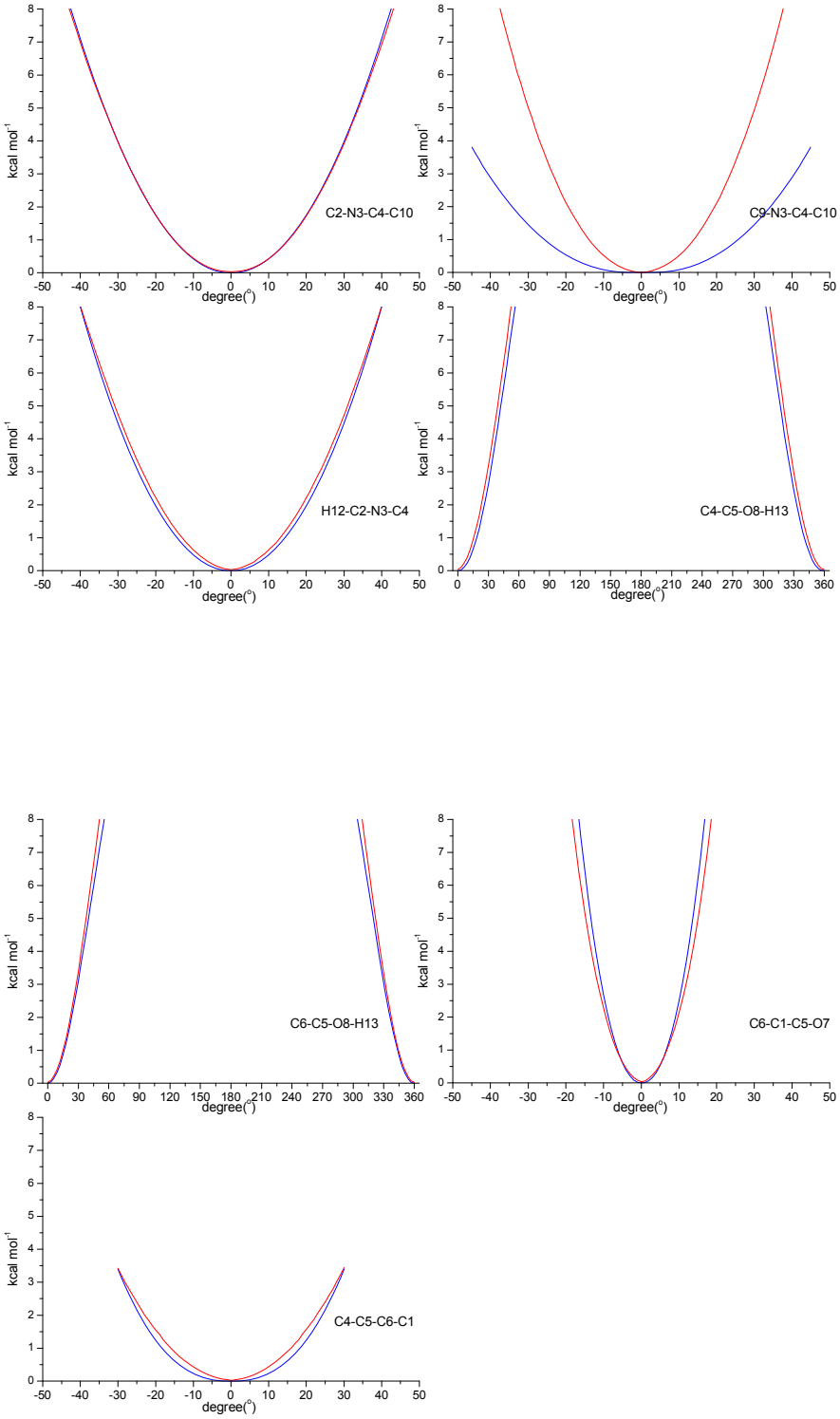
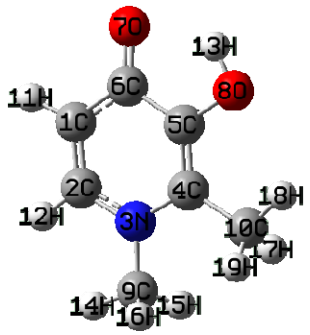


Figure 6-3 continued

Table 6-5 Atom types and charges specific for CP20.



Atom	Atom type	Charge (initial)	Charge (final)
C1	CG2R62	-0.050	-0.480
C2	CG2R62	0.179	-0.310
N3	NG2R61	-0.811	0.065
C4	CG2R62	0.177	0.195
C5	CG2R62	0.161	0.040
C6	CG2R63	0.604	0.635
O7	OG2D4	-0.477	-0.645
O8	OG311	-0.529	-0.550
C9	CG331	-0.267	-0.230
C10	CG331	-0.165	-0.310
H11	HGR62	0.077	0.270
H12	HGR62	0.141	0.340
H13	HGP1	0.420	0.440
H14	HGA3	0.090	0.090
H15	HGA3	0.090	0.090
H16	HGA3	0.090	0.090
H17	HGA3	0.090	0.090
H18	HGA3	0.090	0.090
H19	HGA3	0.090	0.090

Table 6-6 The force field parameters of bonds, angles, dihedrals and improper dihedrals specific for CP20.

Parameter types				Force constant (initial)	Periodicity of dihedrals (initial)	Optimal parameter (initial)	Force constant (final)	Periodicity of dihedrals (final)	Optimal parameter (final)
Bonds									
CG2R62	OG311			334.30		1.411	150.00		1.2900
Angles									
CG2R62	CG2R62	OG311		45.20		120.00	30.00		119.82
CG2R63	CG2R62	OG311		45.20		120.00	30.00		108.68
CG331	CG2R62	NG2R61		45.80		122.30	20.00		135.00
CG2R62	CG2R63	CG2R62		40.00		118.00	30.00		90.26
CG2R62	OG311	HGP1		65.00		108.00	50.00		102.41
Dihedrals									
CG2R63	CG2R62	CG2R62	CG331	3.10	2	180.00	1.50	2	180.00
CG331	CG2R62	CG2R62	OG311	2.40	2	180.00	4.50	2	180.00
NG2R61	CG2R62	CG2R62	OG311	1.00	2	180.00	8.30	2	180.00
CG2R62	CG2R62	CG2R63	CG2R62	6.00	2	180.00	1.80	2	180.00
OG311	CG2R62	CG2R63	CG2R62	3.10	2	180.00	1.80	2	180.00
OG311	CG2R62	CG2R63	OG2D4	1.00	2	180.00	2.00	2	180.00
HGR62	CG2R62	CG2R63	CG2R62	1.00	2	180.00	0.00	1	0.00
NG2R61	CG2R62	CG331	HGA3	0.00	6	180.00	0.65	3	0.00
CG331	CG2R62	NG2R61	CG2R62	4.00	2	180.00	3.50	2	180.00
CG331	CG2R62	NG2R61	CG331	11.00	2	180.00	0.00	1	180.00
HGR62	CG2R62	NG2R61	CG2R62	4.60	2	180.00	5.00	2	180.00
CG2R62	CG2R62	OG311	HGP1	0.99	2	180.00	2.00	2	180.00
CG2R63	CG2R62	OG311	HGP1	0.99	2	180.00	2.00	2	180.00
Improper dihedrals									
CG2R63	CG2R62	CG2R62	OG2D4	90.00	0	0.00	15.00	0	0.00

Table 6-7 CHARMM molecular mechanics (MM) equilibrium geometry of CP25, compared to QM MP2/6-31G(d) level.

Coordinate	MP2	MM	Difference	Coordinate	MP2	MM	Difference
Bond lengths(Å)				Dihedrals(°)			
C1-C2	1.372	1.365	-0.006	H3-C1-C2-H4	0.0	-0.9	-0.9
C1-H3	1.085	1.088	0.003	H3-C1-C2-N11	-178.9	-179.1	-0.3
C1-C18	1.430	1.436	0.006	C18-C1-C2-H4	-179.4	178.2	-2.3
C2-H4	1.087	1.094	0.008	C18-C1-C2-N11	1.7	0.0	-1.6
C2-N11	1.364	1.353	-0.011	C2-C1-C18-C15	-1.8	0.9	2.8
C5-H6	1.095	1.114	0.019	C2-C1-C18-O19	179.4	-179.3	1.3
C5-H7	1.092	1.110	0.018	H3-C1-C18-C15	178.7	-179.9	1.4
C5-C9	1.526	1.537	0.011	H3-C1-C18-O19	0.0	-0.1	-0.1
C5-N11	1.468	1.466	-0.002	C1-C2-N11-C5	-172.9	-172.1	0.8
H8-C12	1.096	1.111	0.016	C1-C2-N11-C14	-1.6	-2.8	-1.2
C9-C20	1.527	1.536	0.009	H4-C2-N11-C5	8.1	9.7	1.5
C9-H24	1.097	1.114	0.017	H4-C2-N11-C14	179.5	178.9	-0.5
C9-H25	1.096	1.113	0.017	H6-C5-C9-C20	-61.2	-60.5	0.8
H10-C12	1.091	1.111	0.020	H6-C5-C9-H24	60.6	60.8	0.1
N11-C14	1.389	1.374	-0.015	H6-C5-C9-H25	176.6	178.2	1.5
C12-H13	1.094	1.110	0.016	H7-C5-C9-C20	57.3	57.4	0.1
C12-C14	1.499	1.501	0.002	H7-C5-C9-H24	179.1	178.7	-0.5
C14-C15	1.375	1.376	0.001	H7-C5-C9-H25	-64.9	-63.9	0.9
C15-O17	1.357	1.359	0.002	N11-C5-C9-C20	179.2	-179.4	1.4
C15-C18	1.454	1.460	0.007	N11-C5-C9-H24	-58.9	-58.2	0.8
H16-O17	0.994	0.970	-0.023	N11-C5-C9-H25	57.1	59.2	2.2
C18-O19	1.260	1.236	-0.025	H6-C5-N11-C2	-35.4	-35.4	0.1
C20-C21	1.527	1.536	0.009	H6-C5-N11-C14	153.3	155.6	2.3
C20-H26	1.099	1.114	0.015	H7-C5-N11-C2	-150.7	-149.8	0.9
C20-H27	1.099	1.113	0.015	H7-C5-N11-C14	38.0	41.1	3.1
C21-C22	1.527	1.535	0.008	C9-C5-N11-C2	86.0	85.4	-0.6
C21-H28	1.098	1.113	0.015	C9-C5-N11-C14	-85.3	-83.7	1.6
C21-H29	1.098	1.113	0.015	C5-C9-C20-C21	179.3	180.0	0.6
C22-C23	1.526	1.530	0.004	C5-C9-C20-H26	-58.8	-58.8	0.0
C22-H30	1.097	1.114	0.017	C5-C9-C20-H27	57.5	58.7	1.2
C22-H31	1.097	1.114	0.017	H24-C9-C20-C21	58.3	58.5	0.2
C23-H32	1.094	1.111	0.018	H24-C9-C20-H26	-179.9	179.7	-0.5
C23-H33	1.094	1.111	0.017	H24-C9-C20-H27	-63.6	-62.8	0.8
C23-H34	1.094	1.111	0.017	H25-C9-C20-C21	-58.9	-58.1	0.7
Angles(°)				H25-C9-C20-H26	63.0	63.1	0.1
C2-C1-H3	119.9	121.3	1.3	H25-C9-C20-H27	179.3	-179.4	1.3
C2-C1-C18	120.4	121.4	1.1	C2-N11-C14-C12	178.2	179.2	1.0
H3-C1-C18	119.7	117.3	-2.4	C2-N11-C14-C15	1.7	4.4	2.7
C1-C2-H4	122.0	120.9	-1.1	C5-N11-C14-C12	-10.7	-12.2	-1.5
C1-C2-N11	122.9	121.9	-0.9	C5-N11-C14-C15	172.8	173.0	0.2
H4-C2-N11	115.2	117.2	2.0	H8-C12-C14-N11	-64.2	-62.4	1.9
H6-C5-H7	107.0	106.4	-0.6	H8-C12-C14-C15	112.1	112.4	0.3
H6-C5-C9	110.4	109.9	-0.5	H10-C12-C14-N11	177.1	178.2	1.1
H6-C5-N11	107.1	106.7	-0.4	H10-C12-C14-C15	-6.5	-7.1	-0.5
H7-C5-C9	111.1	111.8	0.8	H13-C12-C14-N11	57.8	58.6	0.8
H7-C5-N11	108.6	108.9	0.3	H13-C12-C14-C15	-125.9	-126.6	-0.8
C9-C5-N11	112.4	112.8	0.4	N11-C14-C15-O17	178.7	178.1	-0.5
C5-C9-C20	112.0	112.7	0.7	N11-C14-C15-C18	-2.1	-3.4	-1.3
C5-C9-C21	108.7	109.3	0.7	C12-C14-C15-O17	2.2	3.4	1.2
C5-C9-C22	109.3	109.7	0.4	C12-C14-C15-C18	-178.5	-178.1	0.4
C20-C9-C21	110.1	108.9	-1.2	C14-C15-O17-H16	179.0	179.2	0.2
C20-C9-C22	110.0	108.8	-1.2	C18-C15-O17-H16	-0.3	0.7	0.9
C21-C9-C22	106.6	107.3	0.7	C14-C15-C18-C1	2.1	0.8	-1.4
C2-N11-C5	118.1	115.9	-2.2	C14-C15-C18-O19	-179.0	-179.0	0.0
C2-N11-C14	120.6	120.7	0.1	O17-C15-C18-C1	-178.6	179.4	-2.1
C5-N11-C14	120.7	122.5	1.7	O17-C15-C18-O19	0.3	-0.4	-0.7
H8-C12-H10	108.0	107.2	-0.8	C9-C20-C21-C22	179.9	179.9	0.0
H8-C12-H13	108.2	108.8	0.6	C9-C20-C21-H28	-58.1	-58.8	-0.7
H8-C12-C14	111.8	110.5	-1.3	C9-C20-C21-H29	57.9	58.7	0.8
H10-C12-H13	108.2	106.9	-1.3	H26-C20-C21-C22	58.0	58.7	0.7
H10-C12-C14	107.9	112.0	4.1	H26-C20-C21-H28	180.0	180.0	0.0
H13-C12-C14	112.5	111.2	-1.3	H26-C20-C21-H29	-64.0	-62.6	1.5
N11-C14-C12	120.3	119.2	-1.1	H27-C20-C21-C22	-58.1	-58.7	-0.7
N11-C14-C15	117.8	119.8	2.0	H27-C20-C21-H28	63.9	62.6	-1.4
C12-C14-C15	121.8	120.7	-1.0	H27-C20-C21-H29	179.9	-180.0	0.1
C14-C15-O17	121.6	124.0	2.3	C20-C21-C22-C23	-180.0	180.0	0.0
C14-C15-C18	124.0	121.8	-2.2	C20-C21-C22-H30	-57.9	-58.4	-0.5
O17-C15-C18	114.4	114.2	-0.2	C20-C21-C22-H31	57.9	58.4	0.5
C15-O17-H16	101.4	101.3	-0.1	H28-C21-C22-C23	58.0	58.4	0.4
C1-C18-C15	114.3	114.2	-0.1	H28-C21-C22-H30	-179.9	180.0	-0.1
C1-C18-O19	127.6	122.0	-5.6	H28-C21-C22-H31	-64.1	-63.2	0.9
C15-C18-O19	118.1	123.8	5.7	H29-C21-C22-C23	-58.0	-58.4	-0.4
C9-C20-C21	112.7	112.6	-0.1	H29-C21-C22-H30	64.1	63.1	-0.9
C9-C20-H26	109.5	109.1	-0.4	H29-C21-C22-H31	179.9	180.0	0.0
C9-C20-H27	109.5	109.2	-0.4	C21-C22-C23-H32	180.0	180.0	0.0
C21-C20-H26	109.3	109.0	-0.2	C21-C22-C23-H33	-59.8	-60.0	-0.1
C21-C20-H27	109.3	109.1	-0.2	C21-C22-C23-H34	59.8	60.0	0.1
H26-C20-H27	106.4	107.8	1.4	H30-C22-C23-H32	58.2	58.6	0.4

C20-C21-C22	113.1	113.1	-0.1	H30-C22-C23-H33	178.4	178.7	0.3
C20-C21-H28	109.3	109.3	0.0	H30-C22-C23-H34	-61.9	-61.4	0.5
C20-C21-H29	109.3	109.3	0.0	H31-C22-C23-H32	-58.2	-58.6	-0.4
C22-C21-H28	109.3	108.8	-0.5	H31-C22-C23-H33	61.9	61.4	-0.5
C22-C21-H29	109.3	108.7	-0.6	H31-C22-C23-H34	-178.4	-178.6	-0.3
H28-C21-H29	106.3	107.5	1.2	Improper dihedrals(°)			
C21-C22-C23	112.7	113.1	0.4	C18-C1-C15-O19	0.6	-0.1	-0.7
C21-C22-H30	109.2	108.9	-0.3				
C21-C22-H31	109.2	108.9	-0.3				
C23-C22-H30	109.6	109.2	-0.4				
C23-C22-H31	109.6	109.2	-0.4				
H30-C22-H31	106.4	107.4	1.1				
C22-C23-H32	111.4	110.6	-0.8				
C22-C23-H33	110.9	110.5	-0.3				
C22-C23-H34	110.9	110.5	-0.3				
H32-C23-H33	107.9	108.4	0.5				
H32-C23-H34	107.9	108.4	0.5				
H33-C23-H34	107.8	108.3	0.5				

The coordinates highlighted in blue are specific for CP25 and were optimised in this study.

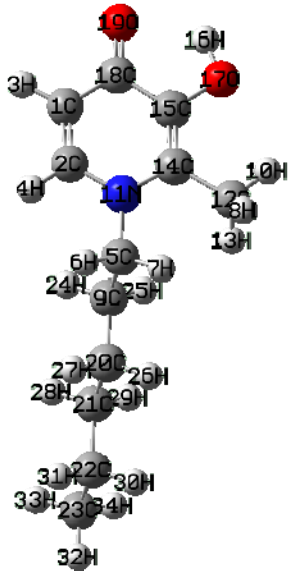


Table 6-8 Interaction energies (kcal mol^{-1}) and distances (Å) of CP25-water complexes in different geometries.

interaction geometry	Δ*E*(HF)[a]	Δ*E*(MM)	ΔΔ*E*	*r*(HF)[b]	*r*(MM)	Δ*r*
O_4···HOH	-6.31	-6.26	0.05	1.99	1.77	-0.22
O_3···HOH	-2.05	-2.03	0.02	2.25	2.07	-0.18
H_5···OHH	-0.81	-0.81	0.00	2.54	2.49	-0.05
H_6···OHH	-3.85	-3.82	0.03	2.36	2.37	0.01
|*M*|			0.03			0.12

[a] HF/6-31G(d) interaction energies are scaled by a factor 1.16 (Vanommeslaeghe *et al.*, 2010); [b] HF interaction distances are not scaled; however, bulk phase hydrogen bonds should be roughly 0.2 shorter than vacuum (Vanommeslaeghe *et al.*, 2010). |*M*|: mean of absolute deviations.

Table 6-9 Dipole moment (Debye) of CP25 calculated at HF/6-31G(d)[a], MP2/6-31G(d) and MM.

μ	HF/6-31G(d)	MP2/6-31G(d)	MM
X	6.5160	7.1907	6.5488
Y	-0.0025	-0.0752	0.0872
Z	2.6133	2.7548	2.9091
total	7.0205	7.7007	7.1664

[a] The HF dipole moment is calculated on the MP2/6-31G(d) optimised geometry.

Table 6-10 CP25 vibrational spectra calculated at scaled MP2/6-31G(d) and MM.

MP2/6-31G(d)[a]	MM	ΔFrequency%[b]
29.3	31.7	8.2%
32.2	33.3	3.4%
44.9	56.4	25.6%
83.3	82.7	-0.7%
95.4	89.3	-6.4%
119.6	112.0	-6.4%
129.9	133.7	2.9%
149.8	149.1	-0.5%
165.3	156.8	-5.1%
178.8	171.7	-4.0%
229.8	234.5	2.0%
245.0	245.3	0.1%
265.9	280.0	5.3%
303.4	313.0	3.2%
306.8	330.0	7.6%
327.6	348.4	6.3%
359.1	386.9	7.7%
410.8	397.6	-3.2%
414.5	432.4	4.3%
447.0	446.7	-0.1%
471.5	461.3	-2.2%
486.8	491.4	0.9%
531.6	519.8	-2.2%
564.6	561.6	-0.5%
574.0	606.1	5.6%
657.8	631.7	-4.0%
660.9	654.0	-1.0%
703.9	723.8	2.8%
720.8	742.0	2.9%
744.2	747.9	0.5%
754.1	759.1	0.7%
777.5	788.5	1.4%
814.9	818.6	0.5%
869.4	859.0	-1.2%
875.4	916.0	4.6%
881.7	925.6	5.0%
972.5	937.2	-3.6%
993.8	971.8	-2.2%
1013.3	1008.0	-0.5%
1015.4	1009.1	-0.6%
1022.9	1017.3	-0.5%
1037.5	1028.0	-0.9%
1042.5	1030.3	-1.2%
1054.7	1047.6	-0.7%
1108.7	1052.1	-5.1%
1128.3	1069.2	-5.2%
1148.9	1102.3	-4.1%
1196.0	1111.6	-7.1%
1208.1	1137.7	-5.8%
1214.9	1144.0	-5.8%
1238.2	1161.8	-6.2%
1247.9	1171.9	-6.1%
1271.9	1194.2	-6.1%
1275.7	1222.2	-4.2%
1285.7	1231.3	-4.2%
1292.4	1258.0	-2.7%
1297.5	1261.4	-2.8%
1330.9	1299.6	-2.4%
1358.6	1344.3	-1.1%
1368.7	1351.3	-1.3%
1373.0	1359.3	-1.0%
1384.9	1405.4	1.5%
1387.3	1413.9	1.9%
1388.9	1421.8	2.4%
1454.1	1422.7	-2.2%
1459.0	1426.0	-2.3%
1461.4	1426.4	-2.4%
1465.4	1428.0	-2.6%
1469.9	1429.1	-2.8%
1471.9	1434.8	-2.5%
1475.9	1443.6	-2.2%
1478.1	1476.2	-0.1%
1484.7	1513.4	1.9%
1487.5	1538.2	3.4%
1540.9	1568.5	1.8%
1610.1	1624.6	0.9%
1627.5	1768.5	8.7%
2901.6	2853.8	-1.6%
2907.3	2887.5	-0.7%
2918.3	2893.4	-0.9%
2924.1	2896.5	-0.9%
2926.1	2901.8	-0.8%

2926.7	2905.0	-0.7%
2941.0	2907.2	-1.1%
2950.7	2911.4	-1.3%
2953.1	2921.5	-1.1%
2969.3	2926.9	-1.4%
2978.9	2934.1	-1.5%
2995.4	2940.3	-1.8%
3010.0	2957.5	-1.7%
3012.2	2958.9	-1.8%
3014.4	2959.5	-1.8%
3040.3	2960.3	-2.6%
3057.0	2995.0	-2.0%
3082.8	2997.9	-2.8%
3240.5	3677.8	13.5%
	$\|M\|$	3.1%

[a] scaled by a factor of 0.943 (Vanommeslaeghe *et al.*, 2010); [b] ΔFrequency% = (MM_frequency - MP2/6-31G(d)_frequency) (MP2/6-31G(d)_frequency)$^{-1}$ x100%; $|M|$: mean of absolute ΔFrequency%.

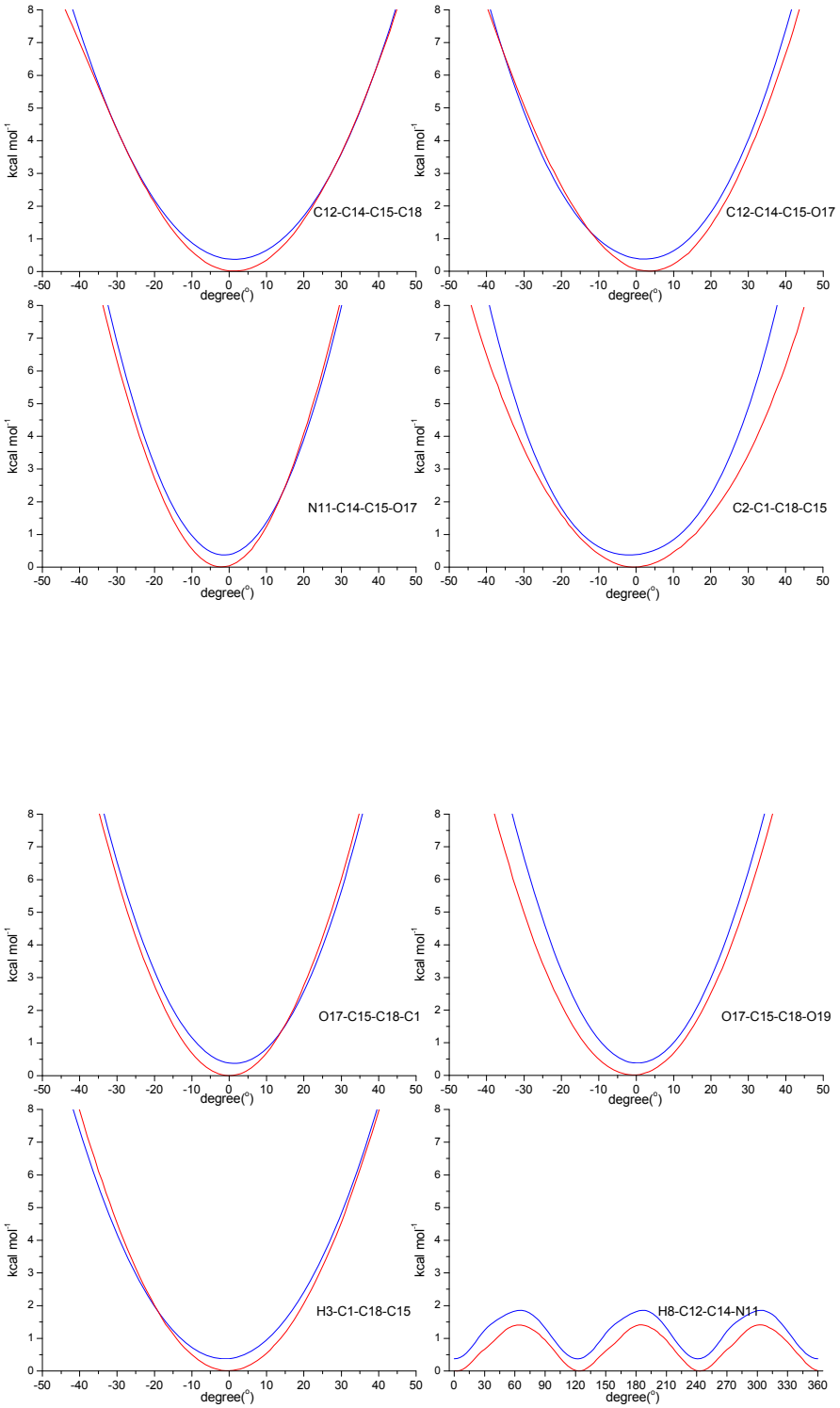


Figure 6-4 Potential energy surfaces of CP25 calculated at MP2/6-31G(d) (blue) and MM (red) for the optimised dihedrals and improper dihedrals.

Y-axis is only shown up to 8 kcal mol^{-1} because this is the only important region for MD simulations at 300 K.

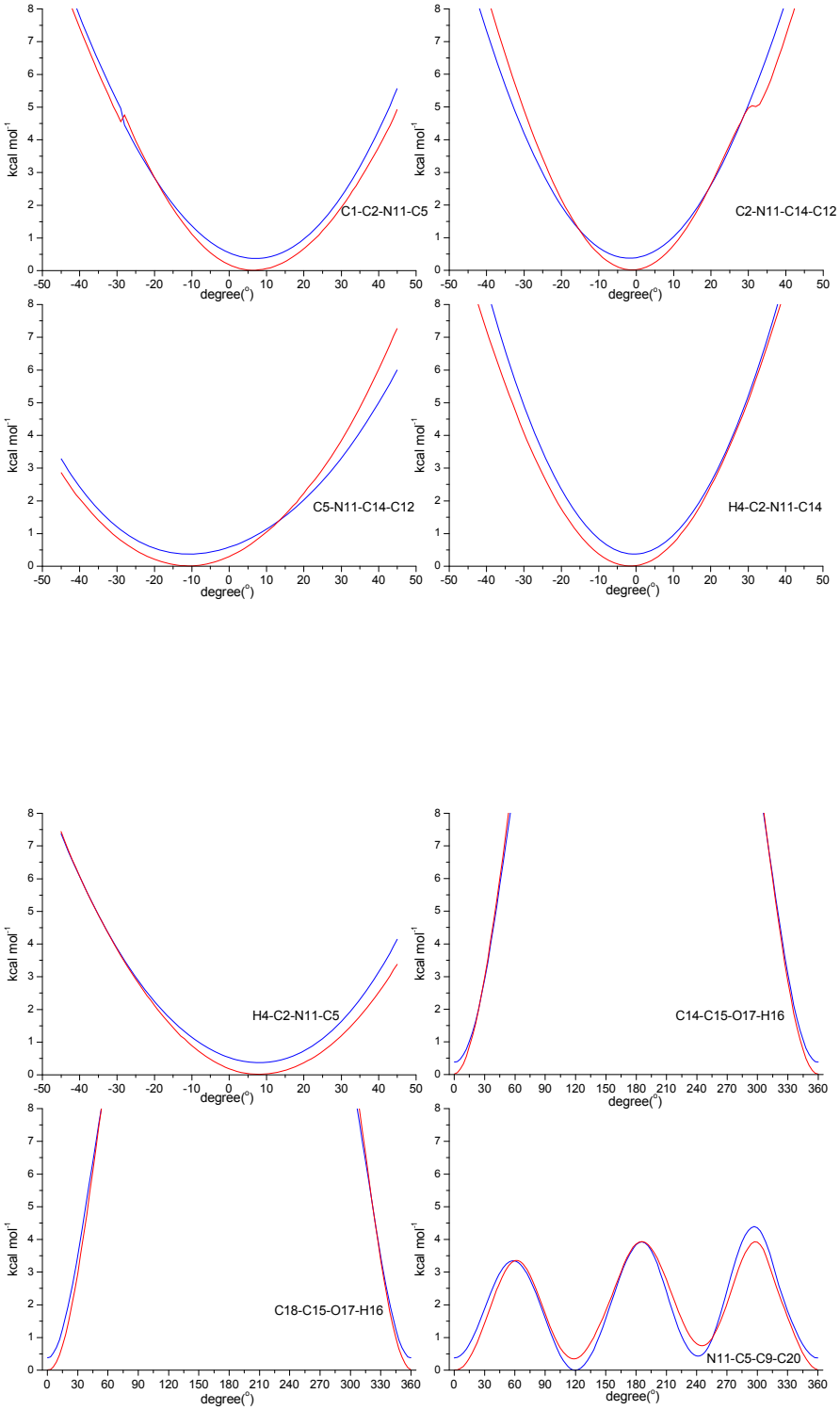


Figure 6-4 continued

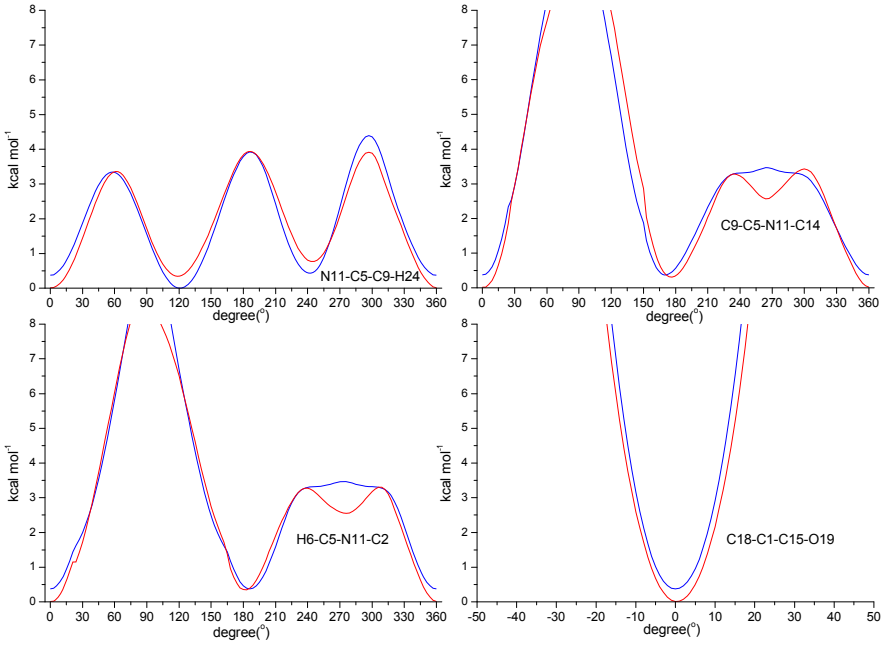
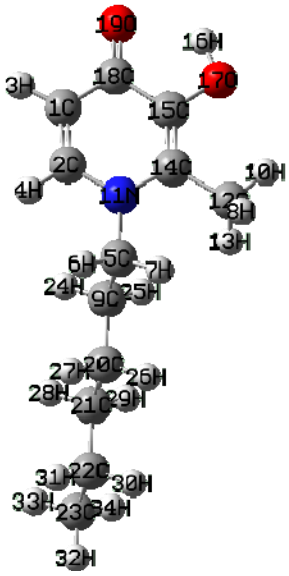


Figure 6-4 continued

Table 6-11 Atom types and charges specific for CP25.



Atom	Atom type	Charge (initial)	Charge (final)
C1	CG2R62	-0.050	-0.406
C2	CG2R62	0.176	-0.129
H3	HGR62	0.077	0.190
H4	HGR62	0.141	0.221
C5	CG321	-0.106	-0.054
H6	HGA2	0.090	0.090
H7	HGA2	0.090	0.090
H8	HGA3	0.090	0.090
C9	CG321	-0.180	-0.137
H10	HGA3	0.090	0.090
N11	NG2R61	-0.861	-0.164
C12	CG331	-0.165	-0.280
H13	HGA3	0.090	0.090
C14	CG2R62	0.174	0.262
C15	CG2R62	0.161	-0.034
H16	HGP1	0.420	0.412
O17	OG311	-0.529	-0.489
C18	CG2R63	0.604	0.493
O19	OG2D4	-0.477	-0.540
C20	CG321	-0.195	-0.196
C21	CG321	-0.178	-0.165
C22	CG321	-0.183	-0.063
C23	CG331	-0.269	-0.361
H24	HGA2	0.090	0.090
H25	HGA2	0.090	0.090
H26	HGA2	0.090	0.090
H27	HGA2	0.090	0.090
H28	HGA2	0.090	0.090
H29	HGA2	0.090	0.090
H30	HGA2	0.090	0.090
H31	HGA2	0.090	0.090
H32	HGA3	0.090	0.090
H33	HGA3	0.090	0.090
H34	HGA3	0.090	0.090

Table 6-12 The force field parameters of bonds, angles, dihedrals and improper dihedrals specific for CP25.

Parameter types				Force constant (initial)	Periodicity of dihedrals (initial)	Optimal parameter (initial)	Force constant (final)	Periodicity of dihedrals (final)	Optimal parameter (final)
Bonds									
CG2R62	OG311			334.30		1.411	150.00		1.300
CG321	NG2R61			400.00		1.456	300.00		1.440
Angles									
CG2R62	CG2R62	OG311		45.20		120.00	30.00		111.30
CG2R63	CG2R62	OG311		45.20		120.00	30.00		100.39
CG331	CG2R62	NG2R61		45.80		122.30	20.00		119.60
CG2R62	CG2R63	CG2R62		40.00		118.00	30.00		89.30
CG321	CG321	NG2R61		70.00		113.50	20.00		107.80
NG2R61	CG321	HGA2		51.50		109.50	40.00		103.50
CG2R62	NG2R61	CG321		70.00		120.50	50.00		108.00
CG2R62	OG311	HGP1		65.00		108.00	50.00		102.50
Dihedrals									
CG2R63	CG2R62	CG2R62	CG331	3.10	2	180.00	1.80	2	180.00
CG331	CG2R62	CG2R62	OG311	2.40	2	180.00	5.00	2	180.00
NG2R61	CG2R62	CG2R62	OG311	1.00	2	180.00	9.00	2	180.00
CG2R62	CG2R62	CG2R63	CG2R62	6.00	2	180.00	2.60	2	180.00
OG311	CG2R62	CG2R63	CG2R62	3.10	2	180.00	1.80	2	180.00
OG311	CG2R62	CG2R63	OG2D4	1.00	2	180.00	2.00	2	180.00
HGR62	CG2R62	CG2R63	CG2R62	1.00	2	180.00	0.01	1	0.00
NG2R61	CG2R62	CG331	HGA3	0.00	6	180.00	0.58	3	12.00
CG2R62	CG2R62	NG2R61	CG321	11.00	2	180.00	3.60	2	190.00
CG331	CG2R62	NG2R61	CG2R62	4.00	2	180.00	4.60	2	180.00
CG331	CG2R62	NG2R61	CG321	11.00	2	180.00	3.00	1	170.00
HGR62	CG2R62	NG2R61	CG2R62	4.60	2	180.00	4.60	2	180.00
HGR62	CG2R62	NG2R61	CG321	0.30	2	180.00	6.00	1	190.00
CG2R62	CG2R62	OG311	HGP1	0.99	2	180.00	2.00	2	180.00
CG2R63	CG2R62	OG311	HGP1	0.99	2	180.00	2.00	2	180.00
CG321	CG321	CG321	NG2R61	0.20	3	0.00	0.20	3	0.00
NG2R61	CG321	CG321	HGA2	0.20	3	0.00	0.00	3	0.00
CG321	CG321	NG2R61	CG2R62	1.80	1	0.00	0.50	2	0.00
HGA2	CG321	NG2R61	CG2R62	0.00	3	0.00	0.00	4	0.00
Improper dihedrals									
CG2R63	CG2R62	CG2R62	OG2D4	90.00	0	0.00	15.00	0	0.00

Table 6-13 CHARMM molecular mechanics (MM) equilibrium geometry of CP94, compared to QM MP2/6-31G(d) level.

Coordinate	QM	MM	Difference	Coordinate	QM	MM	Difference
Bond lengths(Å)				Dihedrals(°)			
C1-C2	1.372	1.362	-0.010	H3-C1-C2-H4	0.1	0.0	-0.1
C1-H3	1.085	1.089	0.004	H3-C1-C2-N11	-178.9	-179.4	-0.5
C1-C18	1.431	1.434	0.004	C18-C1-C2-H4	-179.6	-179.7	-0.2
C2-H4	1.086	1.091	0.004	C18-C1-C2-N11	1.5	0.9	-0.6
C2-N11	1.364	1.358	-0.006	C2-C1-C18-C15	-1.9	0.2	2.1
C5-H6	1.091	1.110	0.019	C2-C1-C18-O19	179.6	-180.0	0.5
C5-H7	1.093	1.114	0.020	H3-C1-C18-C15	178.5	-179.5	2.0
C5-C9	1.523	1.533	0.010	H3-C1-C18-O19	0.0	0.3	0.4
C5-N11	1.469	1.471	0.002	C1-C2-N11-C5	-172.8	-171.5	1.4
H8-C12	1.094	1.109	0.015	C1-C2-N11-C14	-2.0	-2.3	-0.2
C9-H20	1.093	1.112	0.019	H4-C2-N11-C5	8.2	9.2	1.0
C9-H21	1.092	1.111	0.019	H4-C2-N11-C14	179.0	178.3	-0.6
C9-H22	1.094	1.111	0.016	H6-C5-C9-H20	178.1	178.3	0.2
H10-C12	1.094	1.112	0.019	H6-C5-C9-H21	-62.0	-61.8	0.1
N11-C14	1.390	1.381	-0.009	H6-C5-C9-H22	58.3	58.3	0.0
C12-C13	1.533	1.535	0.002	H7-C5-C9-H20	59.2	59.6	0.4
C12-C14	1.501	1.500	-0.001	H7-C5-C9-H21	179.2	179.5	0.3
C13-H23	1.094	1.110	0.016	H7-C5-C9-H22	-60.6	-60.4	0.2
C13-H24	1.093	1.112	0.020	N11-C5-C9-H20	-59.8	-59.8	0.1
C13-H25	1.094	1.111	0.016	N11-C5-C9-H21	60.1	60.1	0.0
C14-C15	1.375	1.381	0.006	N11-C5-C9-H22	-179.6	-179.7	-0.1
C15-O17	1.357	1.356	-0.001	H6-C5-N11-C2	-147.9	-149.3	-1.4
C15-C18	1.452	1.458	0.006	H6-C5-N11-C14	41.4	41.9	0.5
H16-O17	0.994	0.967	-0.027	H7-C5-N11-C2	-33.0	-34.8	-1.8
C18-O19	1.260	1.237	-0.023	H7-C5-N11-C14	156.3	156.4	0.1
Angles(°)				C9-C5-N11-C2	88.3	86.6	-1.8
C2-C1-H3	119.9	119.7	-0.2	C9-C5-N11-C14	-82.4	-82.2	0.2
C2-C1-C18	120.4	120.7	0.2	C2-N11-C14-C12	176.3	177.5	1.2
H3-C1-C18	119.7	119.6	-0.1	C2-N11-C14-C15	3.0	2.5	-0.5
C1-C2-H4	121.8	119.8	-2.0	C5-N11-C14-C12	-13.2	-14.1	-0.9
C1-C2-N11	123.0	123.4	0.4	C5-N11-C14-C15	173.5	170.9	-2.6
H4-C2-N11	115.2	116.8	1.6	H8-C12-C13-H23	-64.3	-62.5	1.8
H6-C5-H7	106.9	106.8	-0.1	H8-C12-C13-H24	176.0	177.5	1.5
H6-C5-C9	111.6	111.8	0.2	H8-C12-C13-H25	56.4	58.0	1.6
H6-C5-N11	108.6	107.9	-0.7	H10-C12-C13-H23	178.8	-180.0	1.3
H7-C5-C9	110.6	110.3	-0.4	H10-C12-C13-H24	59.0	60.0	1.0
H7-C5-N11	106.7	106.8	0.1	H10-C12-C13-H25	-60.6	-59.5	1.1
C9-C5-N11	112.1	112.9	0.8	C14-C12-C13-H23	60.9	60.1	-0.7
C5-C9-H20	110.2	110.6	0.3	C14-C12-C13-H24	-58.8	-59.9	-1.0
C5-C9-H21	111.0	110.8	-0.2	C14-C12-C13-H25	-178.5	-179.4	-0.9
C5-C9-H22	110.2	110.5	0.3	H8-C12-C14-N11	40.2	37.6	-2.6
H20-C9-H21	108.2	108.1	-0.1	H8-C12-C14-C15	-146.7	-147.4	-0.7
H20-C9-H22	108.5	108.3	-0.2	H10-C12-C14-N11	156.0	153.7	-2.3
H21-C9-H22	108.5	108.4	-0.1	H10-C12-C14-C15	-30.9	-31.3	-0.4
C2-N11-C5	117.9	116.2	-1.7	C13-C12-C14-N11	-84.6	-85.6	-1.0
C2-N11-C14	120.4	119.8	-0.6	C13-C12-C14-C15	88.5	89.4	0.9
C5-N11-C14	121.0	123.0	2.0	N11-C14-C15-O17	177.4	178.1	0.7
H8-C12-H10	106.5	106.2	-0.4	N11-C14-C15-C18	-3.7	-1.5	2.2
H8-C12-C13	110.7	111.3	0.6	N12-C14-C15-O17	4.0	3.0	-1.1
H8-C12-C14	111.4	110.4	-1.0	N12-C14-C15-C18	-177.0	-176.6	0.5
H10-C12-C13	109.2	110.1	1.0	C14-C15-O17-H16	178.0	-179.7	2.4
H10-C12-C14	106.5	108.8	2.3	C18-C15-O17-H16	-1.1	-0.1	1.0
C13-C12-C14	112.2	109.9	-2.3	C14-C15-C18-C1	3.1	0.1	-3.0
C12-C13-H23	111.6	111.1	-0.5	C14-C15-C18-O19	-178.2	-179.7	-1.5
C12-C13-H24	109.6	110.2	0.6	C17-C15-C18-C1	-177.9	-179.4	-1.5
C12-C13-H25	110.8	110.4	-0.4	C17-C15-C18-O19	0.8	0.7	-0.1
H23-C13-H24	108.1	108.2	0.2	Improper dihedrals(°)			
H23-C13-H25	108.2	108.6	0.3	C18-C1-C15-O19			
H24-C13-H25	108.4	108.2	-0.2				
N11-C14-C12	121.3	121.4	0.1				
N11-C14-C15	117.7	119.2	1.5				
C12-C14-C15	120.6	119.2	-1.4				
C14-C15-O17	121.1	122.8	1.7				
C14-C15-C18	124.3	122.8	-1.5				
O17-C15-C18	114.6	114.4	-0.2				
C15-O17-H16	101.4	101.5	0.1				
C1-C18-C15	114.0	114.1	0.1				
C1-C18-O19	127.8	122.5	-5.3				
C15-C18-O19	118.2	123.4	5.2				

The coordinates highlighted in blue are specific for CP94 and were optimised in this study.

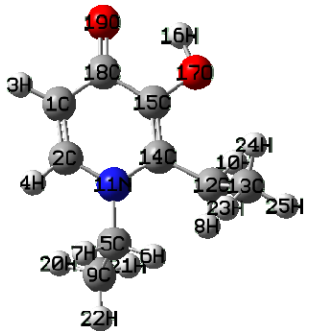


Table 6-14 Interaction energies (kcal mol^{-1}) and distances (Å) of CP94-water complexes in different geometries.

interaction geometry	Δ*E*(HF)[a]	Δ*E*(MM)	ΔΔ*E*	*r*(HF)[b]	*r*(MM)	Δ*r*
O_4···HOH	-5.02	-5.09	-0.07	1.99	1.81	-0.18
O_3···HOH	-0.89	-0.83	0.06	2.23	2.22	-0.01
H_5···OHH	0.44	0.53	0.09	2.53	2.86	0.33
H_6···OHH	-2.76	-2.69	0.07	2.35	2.44	0.09
|*M*|			0.07			0.15

[a] HF/6-31G(d) interaction energies are scaled by a factor 1.16 (Vanommeslaeghe *et al.*, 2010); [b] HF interaction distances are not scaled; however, bulk phase hydrogen bonds should be roughly 0.2 shorter than vacuum (Vanommeslaeghe *et al.*, 2010). |*M*|: mean of absolute deviations.

Table 6-15 Dipole moment (Debye) of CP94 calculated at HF/6-31G(d)[a], MP2/6-31G(d) and MM.

μ	HF/6-31G(d)	MP2/6-31G(d)	MM
X	-6.4330	-7.1254	-6.1167
Y	-1.1340	-1.3571	0.5134
Z	-0.7036	-0.8248	-0.1523
total	6.5700	7.3002	6.1401

[a] The HF dipole moment is calculated on the MP2/6-31G(d) optimised geometry.

Table 6-16 CP94 vibrational spectra calculated at scaled MP2/6-31G(d) and MM.

MP2/6-31G(d)[a]	MM	ΔFrequency%[b]
57.8	65.9	14.0%
76.8	88.5	15.2%
95.6	108.3	13.3%
113.1	113.2	0.1%
160.8	161.7	0.6%
206.6	209.3	1.3%
211.2	217.8	3.1%
236.8	224.9	-5.0%
307.6	291.8	-5.1%
330.7	306.3	-7.4%
335.0	318.0	-5.1%
364.2	341.0	-6.4%
385.2	381.7	-0.9%
455.5	412.4	-9.5%
473.5	467.3	-1.3%
516.0	472.0	-8.5%
542.1	503.3	-7.2%
568.8	516.7	-9.2%
577.2	591.3	2.4%
644.7	639.5	-0.8%
665.7	667.0	0.2%
719.0	711.0	-1.1%
744.9	723.6	-2.9%
759.6	785.1	3.4%
776.6	839.4	8.1%
815.4	857.9	5.2%
874.4	883.4	1.0%
931.7	962.1	3.3%
962.0	972.8	1.1%
1018.2	984.2	-3.3%
1055.0	1026.3	-2.7%
1069.5	1028.2	-3.9%
1078.9	1035.5	-4.0%

1122.2	1040.4	-7.3%
1146.1	1068.6	-6.8%
1206.1	1136.9	-5.7%
1229.3	1139.0	-7.3%
1247.4	1160.6	-7.0%
1268.3	1195.2	-5.8%
1282.4	1276.5	-0.5%
1328.1	1322.0	-0.5%
1339.8	1331.9	-0.6%
1367.5	1399.8	2.4%
1379.4	1416.3	2.7%
1384.9	1421.0	2.6%
1391.1	1423.3	2.3%
1455.2	1423.8	-2.2%
1462.5	1425.7	-2.5%
1468.3	1431.0	-2.5%
1472.3	1431.9	-2.7%
1475.8	1473.8	-0.1%
1481.6	1503.0	1.4%
1488.5	1525.0	2.5%
1540.0	1555.9	1.0%
1604.0	1617.4	0.8%
1627.7	1755.9	7.9%
2928.4	2853.0	-2.6%
2935.9	2855.0	-2.8%
2948.7	2886.5	-2.1%
2964.6	2890.1	-2.5%
2994.9	2903.4	-3.1%
3010.9	2903.7	-3.6%
3017.9	2958.3	-2.0%
3024.2	2958.5	-2.2%
3026.4	2960.3	-2.2%
3037.4	2960.5	-2.5%
3057.2	2992.1	-2.1%
3082.2	2997.9	-2.7%
3241.3	3678.4	13.5%
	$\|M\|$	4.0%

[a] scaled by a factor of 0.943 (Vanommeslaeghe *et al.*, 2010); [b] ΔFrequency% = (MM_frequency - MP2/6-31G(d)_frequency) (MP2/6-31G(d)_frequency)$^{-1}$ x100%; $|M|$: mean of absolute ΔFrequency%.

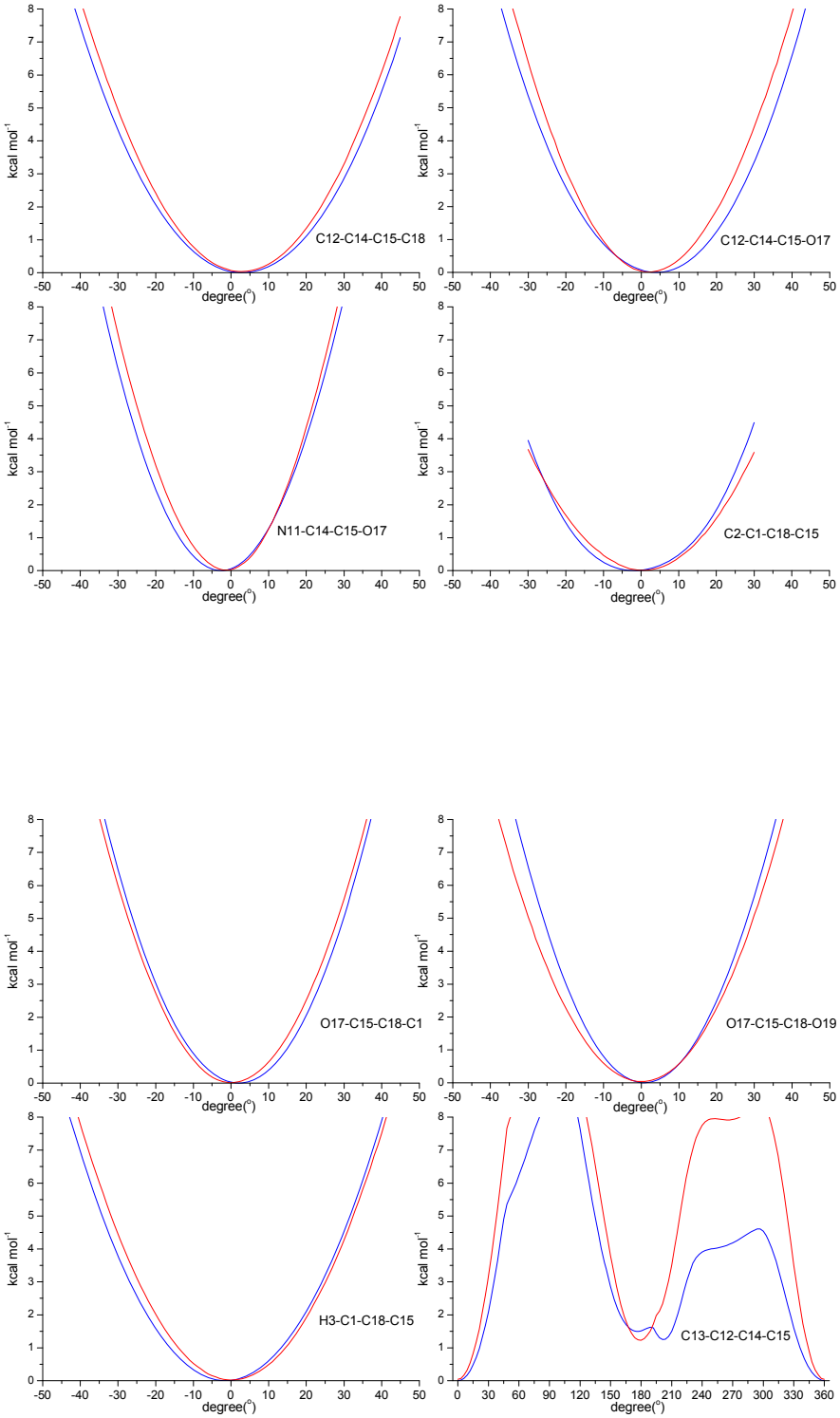


Figure 6-5 Potential energy surfaces of CP94 calculated at MP2/6-31G(d) (blue) and MM (red) for the optimised dihedrals and improper dihedrals.

Y-axis is only shown up to 8 kcal mol^{-1} because this is the only important region for MD simulations at 300 K.

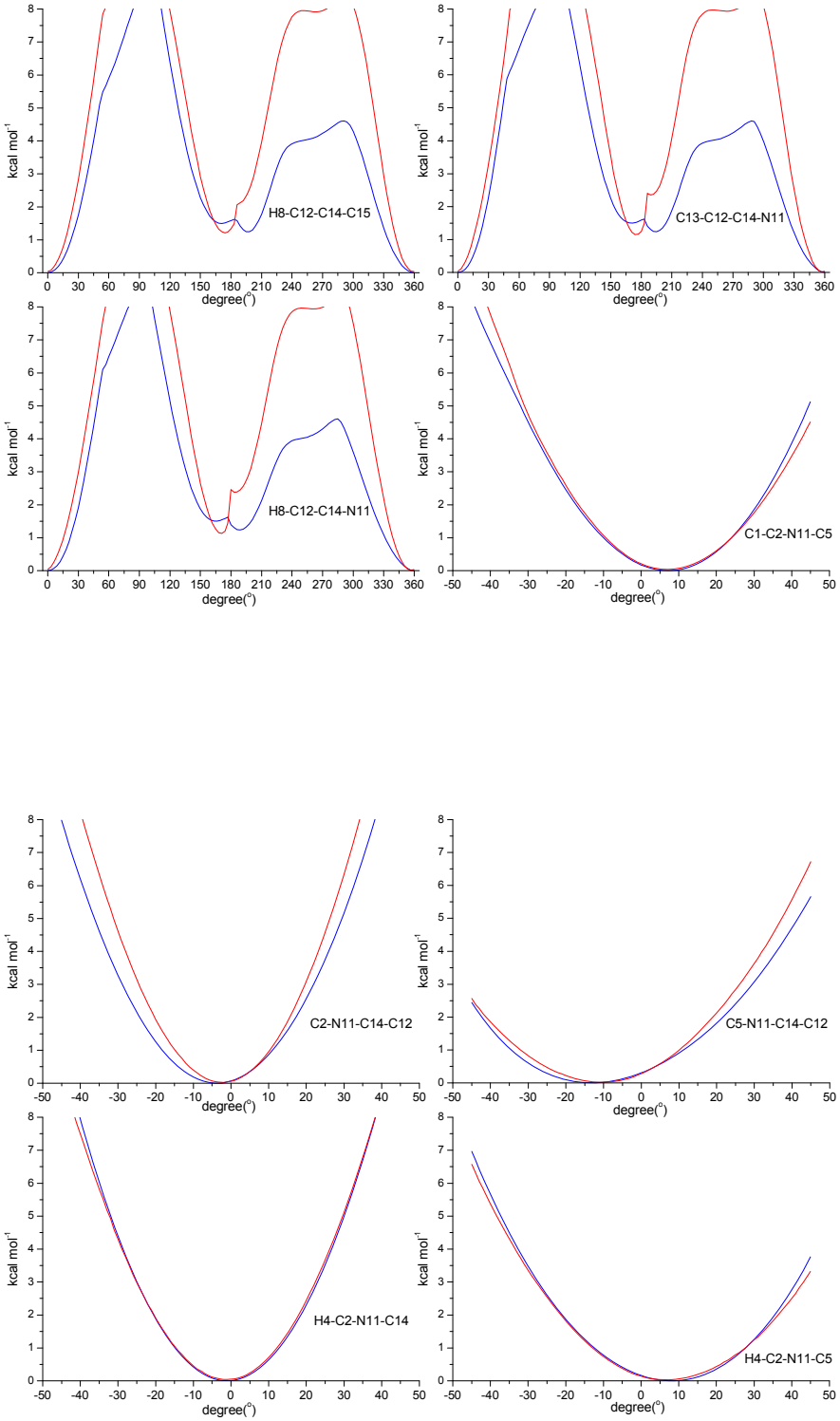


Figure 6-5 continued

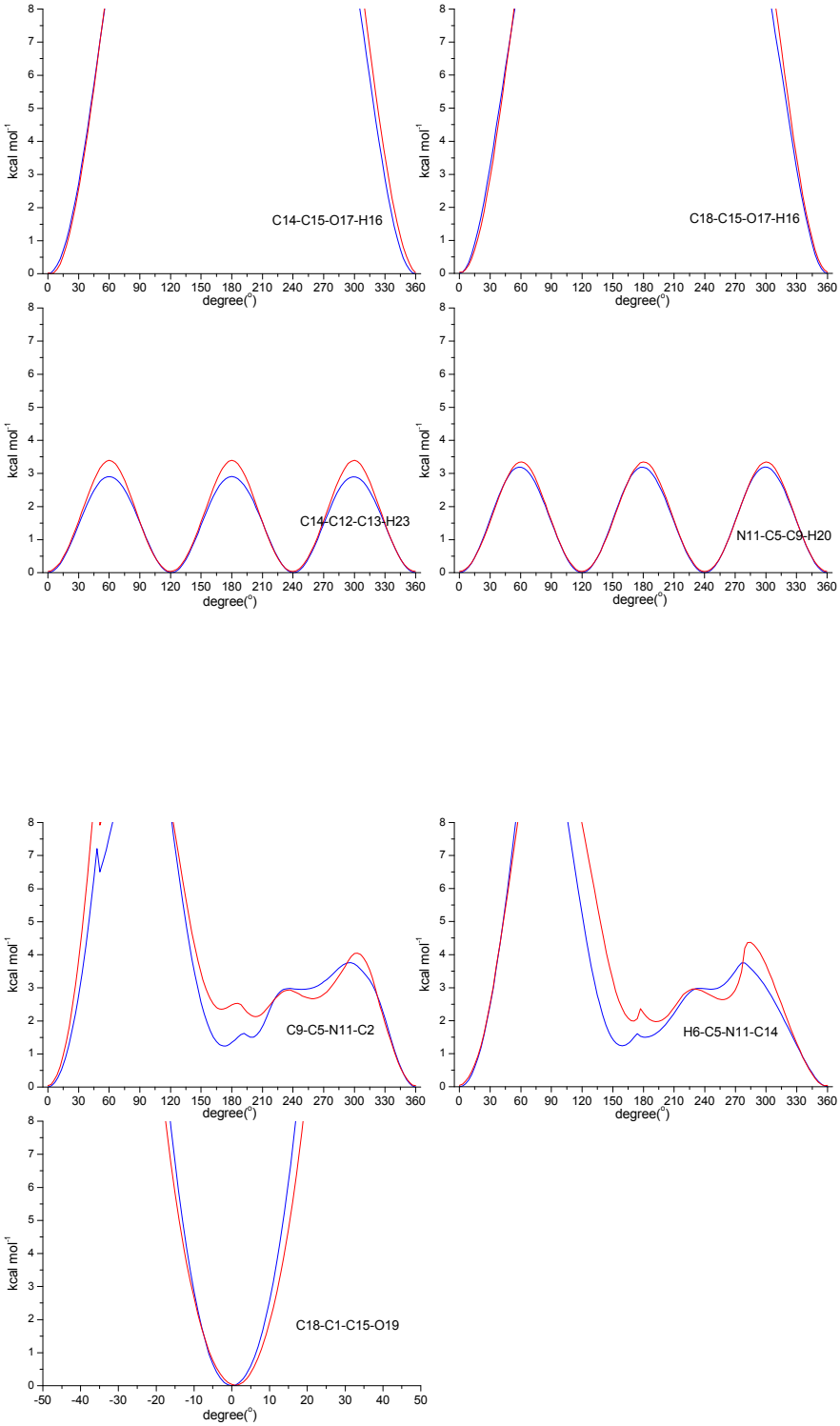
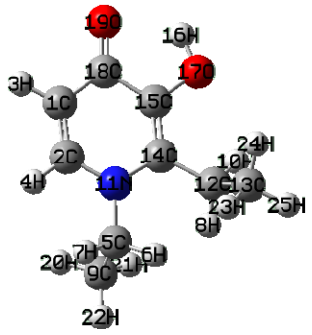


Figure 6-5 continued

Table 6-17 Atom types and charges specific for CP94.



Atom	Atom type	Charge (initial)	Charge (final)
C1	CG2R62	-0.050	-0.272
C2	CG2R62	0.176	-0.084
H3	HGR62	0.077	0.072
H4	HGR62	0.141	0.145
C5	CG321	-0.109	0.020
H6	HGA2	0.090	0.090
H7	HGA2	0.090	0.090
H8	HGA2	0.090	0.090
C9	CG331	-0.269	-0.269
H10	HGA2	0.090	0.090
N11	NG2R61	-0.898	-0.052
C12	CG321	-0.172	-0.048
C13	CG331	-0.267	-0.283
C14	CG2R62	0.301	-0.035
C15	CG2R62	0.148	-0.092
H16	HGP1	0.420	0.402
O17	OG311	-0.529	-0.349
C18	CG2R63	0.608	0.388
O19	OG2D4	-0.477	-0.443
H20	HGA3	0.090	0.090
H21	HGA3	0.090	0.090
H22	HGA3	0.090	0.090
H23	HGA3	0.090	0.090
H24	HGA3	0.090	0.090
H25	HGA3	0.090	0.090

Table 6-18 The force field parameters of bonds, angles, dihedrals and improper dihedrals specific for CP94.

Parameter types				Force constant (initial)	Periodicity of dihedrals (initial)	Optimal parameter (initial)	Force constant (final)	Periodicity of dihedrals (final)	Optimal parameter (final)
Bonds									
CG2R62	CG321			230.00		1.490	150.00		1.45
CG2R62	OG311			334.30		1.411	110.00		1.31
CG321	NG2R61			400.00		1.456	300.00		1.45
Angles									
CG2R62	CG2R62	CG321		40.00		124.20	10.00		150.90
CG2R62	CG2R62	OG311		45.20		120.00	20.00		104.80
CG2R63	CG2R62	OG311		45.20		120.00	20.00		99.00
CG321	CG2R62	NG2R61		45.80		122.30	10.00		135.00
CG2R62	CG2R63	CG2R62		40.00		118.00	20.00		83.00
CG2R62	CG321	CG331		51.80		107.50	10.00		100.50
CG2R62	CG321	HGA2		49.30		107.50	30.00		105.50
CG331	CG321	NG2R61		70.00		120.00	20.00		108.50
NG2R61	CG321	HGA2		51.50		109.50	40.00		104.00
CG2R62	NG2R61	CG321		70.00		120.50	50.00		114.00
CG2R62	OG311	HGP1		65.00		108.00	60.00		103.10
Dihedrals									
CG2R63	CG2R62	CG2R62	CG321	3.10	2	180.00	2.00	2	180.00
CG321	CG2R62	CG2R62	OG311	2.40	2	180.00	5.00	2	180.00
NG2R61	CG2R62	CG2R62	OG311	1.00	2	180.00	9.00	2	180.00
CG2R62	CG2R62	CG2R63	CG2R62	6.00	2	180.00	2.80	2	185.00
OG311	CG2R62	CG2R63	CG2R62	3.10	2	180.00	1.80	2	180.00
OG311	CG2R62	CG2R63	OG2D4	1.00	2	180.00	2.00	2	180.00
HGR62	CG2R62	CG2R63	CG2R62	1.00	2	180.00	0.01	1	0.00
CG2R62	CG2R62	CG321	CG331	0.23	2	180.00	0.75	2	0.00
CG2R62	CG2R62	CG321	HGA2	0.00	6	0.00	0.00	4	0.00
NG2R61	CG2R62	CG321	CG331	0.23	2	180.00	0.75	2	0.00
NG2R61	CG2R62	CG321	HGA2	0.00	6	0.00	0.00	4	0.00
CG2R62	CG2R62	NG2R61	CG321	11.00	2	180.00	2.80	2	180.00
CG321	CG2R62	NG2R61	CG2R62	4.00	2	180.00	4.60	2	180.00
CG321	CG2R62	NG2R61	CG321	11.00	2	180.00	2.80	1	110.00
HGR62	CG2R62	NG2R61	CG2R62	4.60	2	180.00	5.00	2	180.00
HGR62	CG2R62	NG2R61	CG321	0.30	2	180.00	2.80	1	180.00
CG2R62	CG2R62	OG311	HGP1	0.99	2	180.00	2.30	2	180.00
CG2R63	CG2R62	OG311	HGP1	0.99	2	180.00	2.30	2	180.00
CG2R62	CG321	CG331	HGA3	0.04	3	0.00	0.12	3	0.00
NG2R61	CG321	CG331	HGA3	0.20	3	0.00	0.17	3	0.00
CG331	CG321	NG2R61	CG2R62	1.80	1	0.00	0.50	2	0.00
HGA2	CG321	NG2R61	CG2R62	0.00	3	0.00	0.00	4	0.00
Improper dihedrals									
CG2R63	CG2R62	CG2R62	OG2D4	90.00	0	0.00	10.00	0	0.00

6.3.2 Hydration study using LAMMPS

The MD simulations of the hydrated HPOs were carried out for three successive 1 ns periods and the corresponding O_3-O (H_2O) and O_4-O (H_2O) RDF results are presented in Figures 6-6 (CP20, O_3), 6-7 (CP20, O_4), 6-8 (CP25, O_3), 6-9 (CP25, O_4), 6-10 (CP94, O_3) and 6-11 (CP94, O_4). There are no significant differences for each 1 ns simulation and thus, the final RDF results were averaged over the full 3 ns period and they are presented in Figures 6-12 (O_3) and 6-13 (O_4). In general, CP20 was found to have the strongest hydrogen bonding interaction with H_2O, followed by CP25 and CP94 (the weakest) and HPO O_4 was found to have stronger hydrogen bonding interaction with H_2O than HPO O_3 with respect to the height and position of RDF, g(r), peaks and the corresponding H_2O coordination numbers.

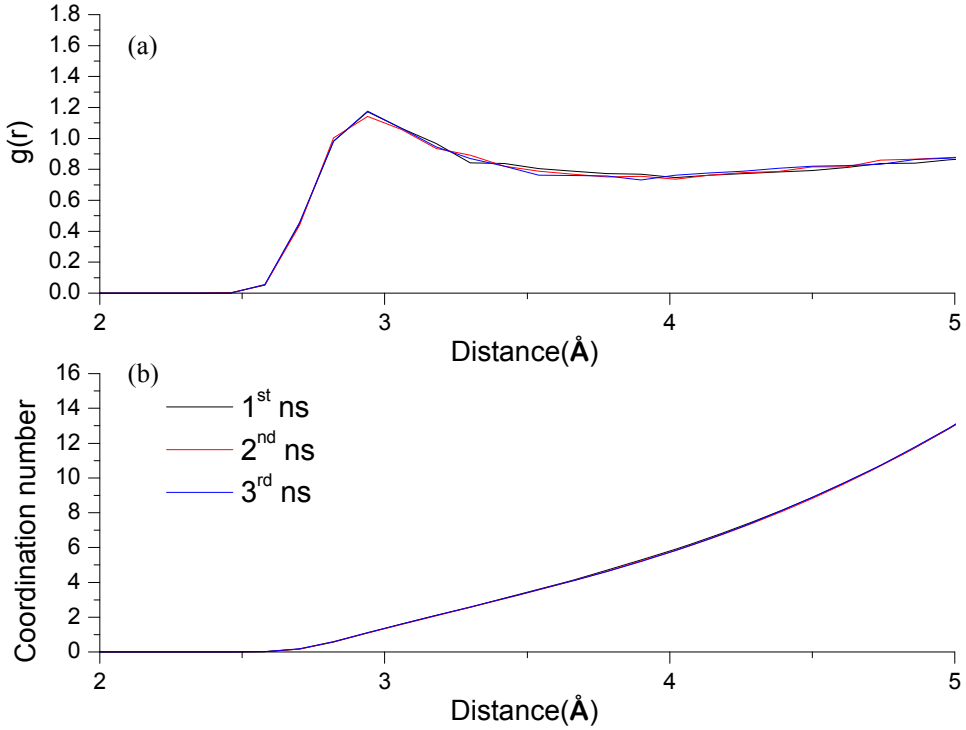


Figure 6-6 O_3–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP20 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

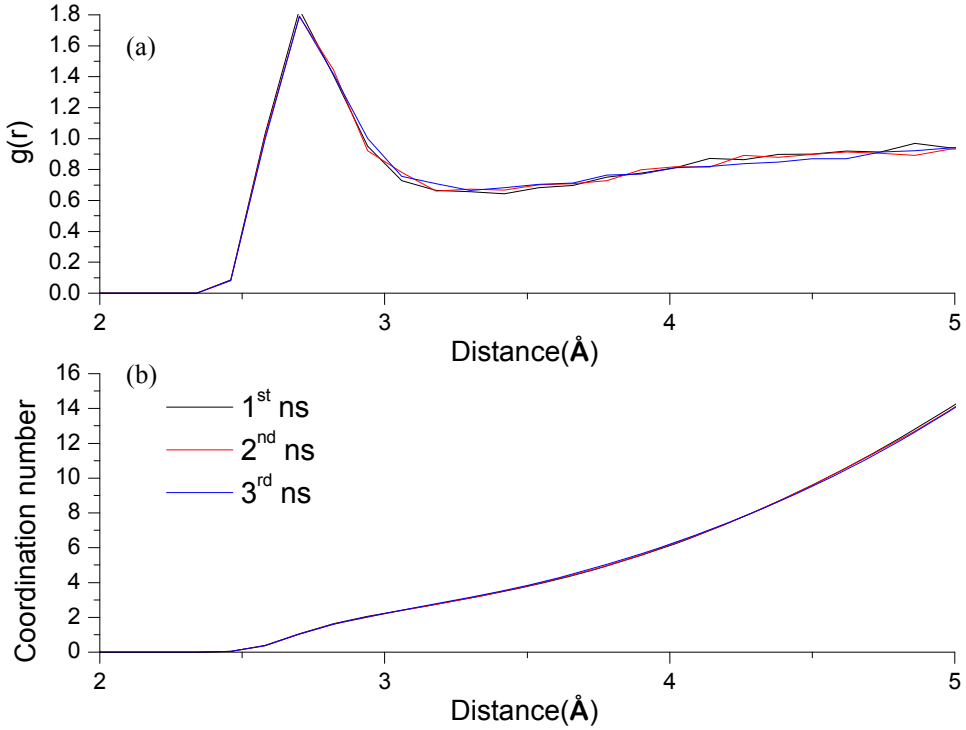


Figure 6-7 O_4–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP20 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

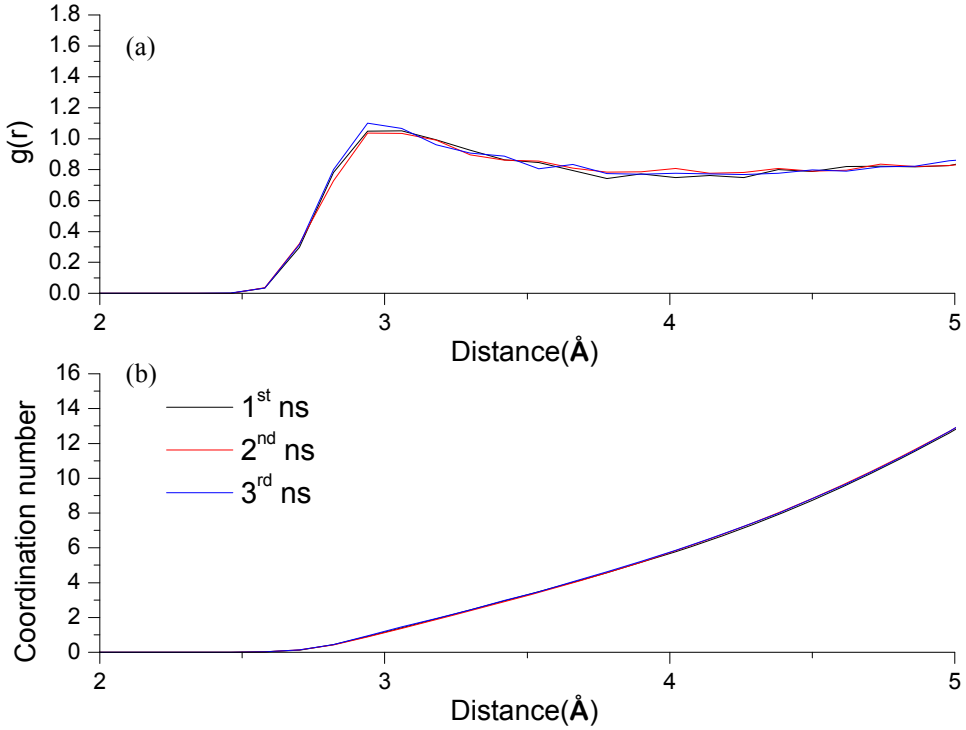


Figure 6-8 O_3–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP25 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

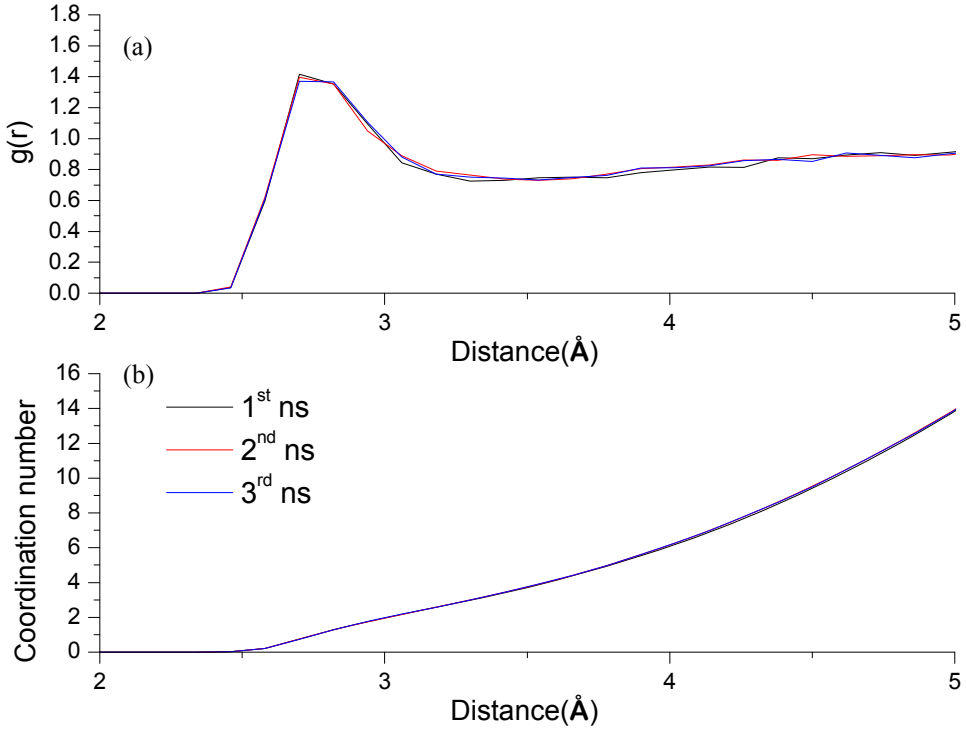


Figure 6-9 O_4–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP25 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

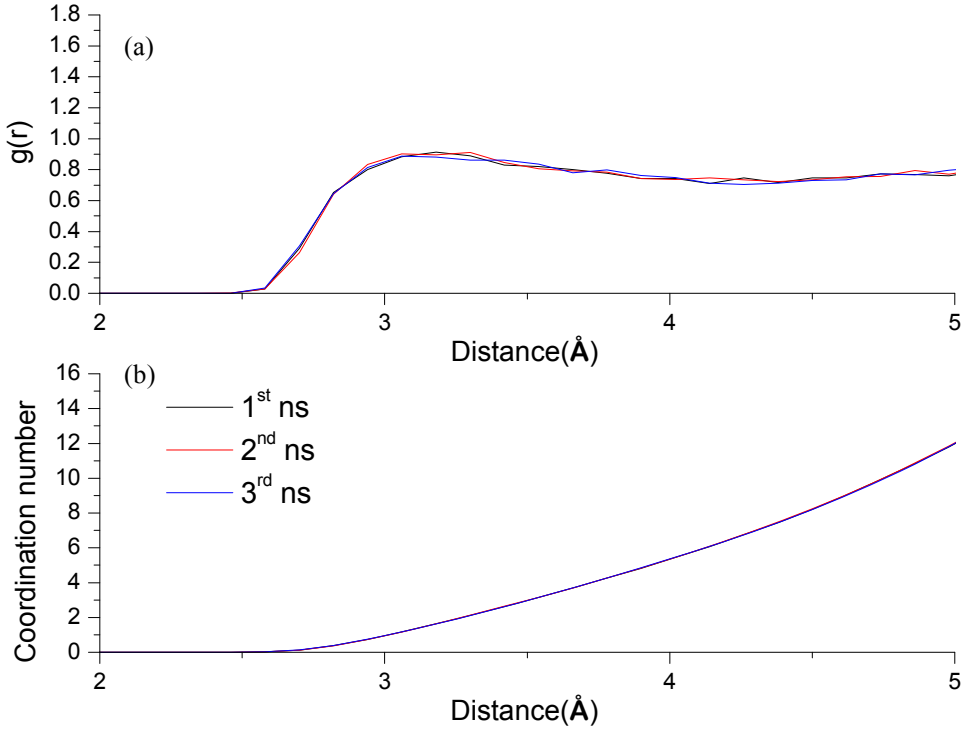


Figure 6-10 O_3–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP94 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

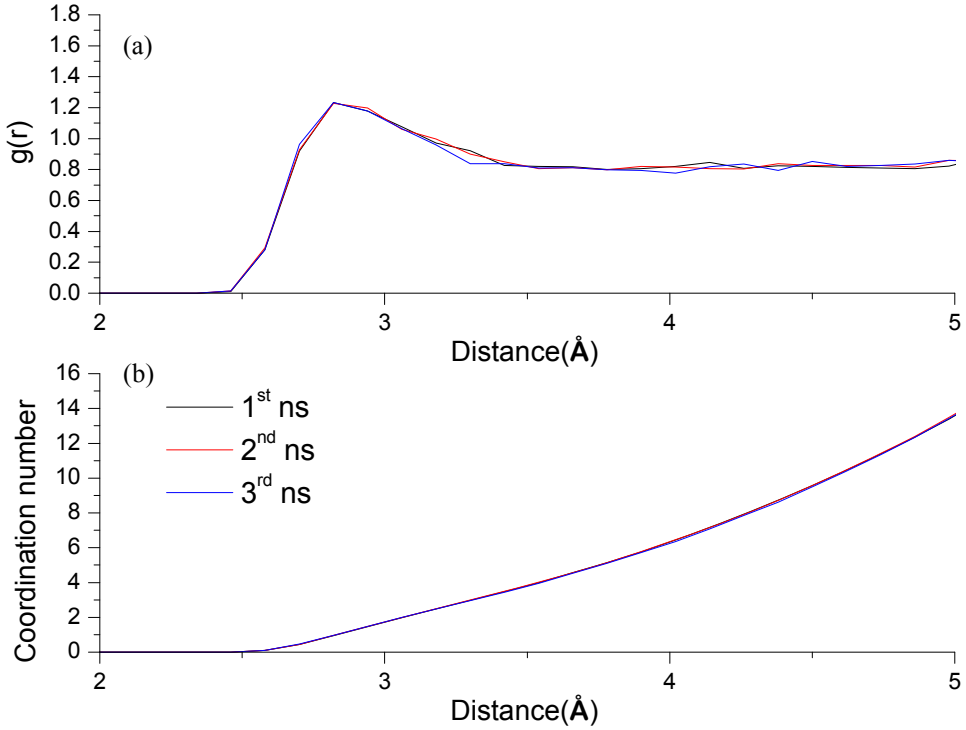


Figure 6-11 O_4–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated CP94 during three successive 1 ns periods[a].

[a] 1st ns (black line), 2nd ns (red line), 3rd ns (blue line).

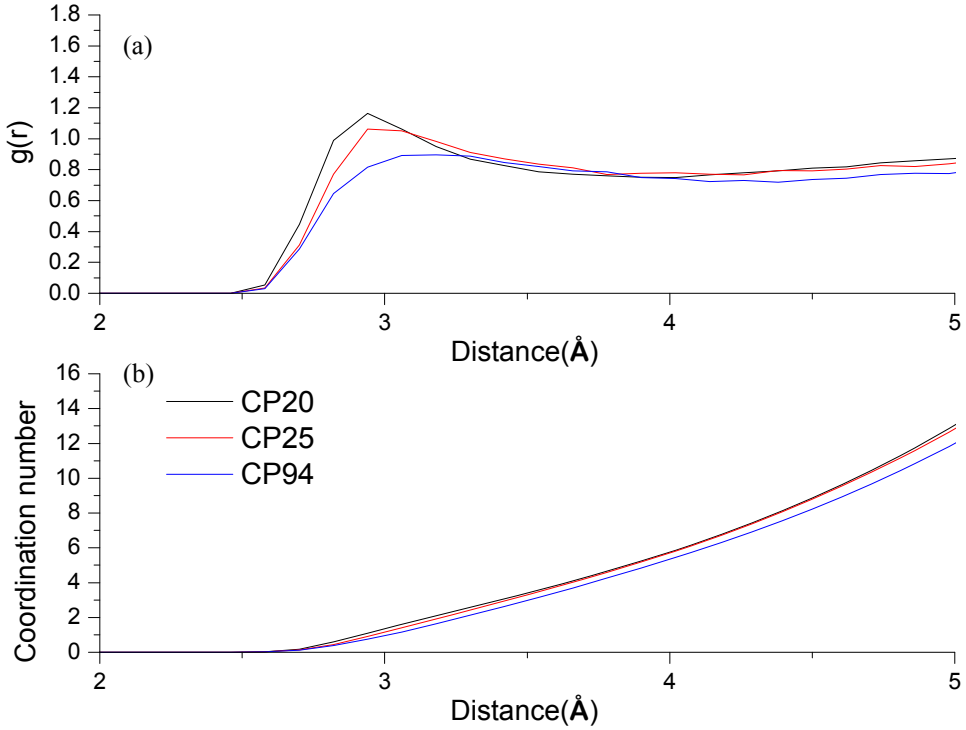


Figure 6-12 O_3–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated HPOs[a] during the whole 3 ns period.

[a] CP20 (black line), CP25 (red line), CP94 (blue line).

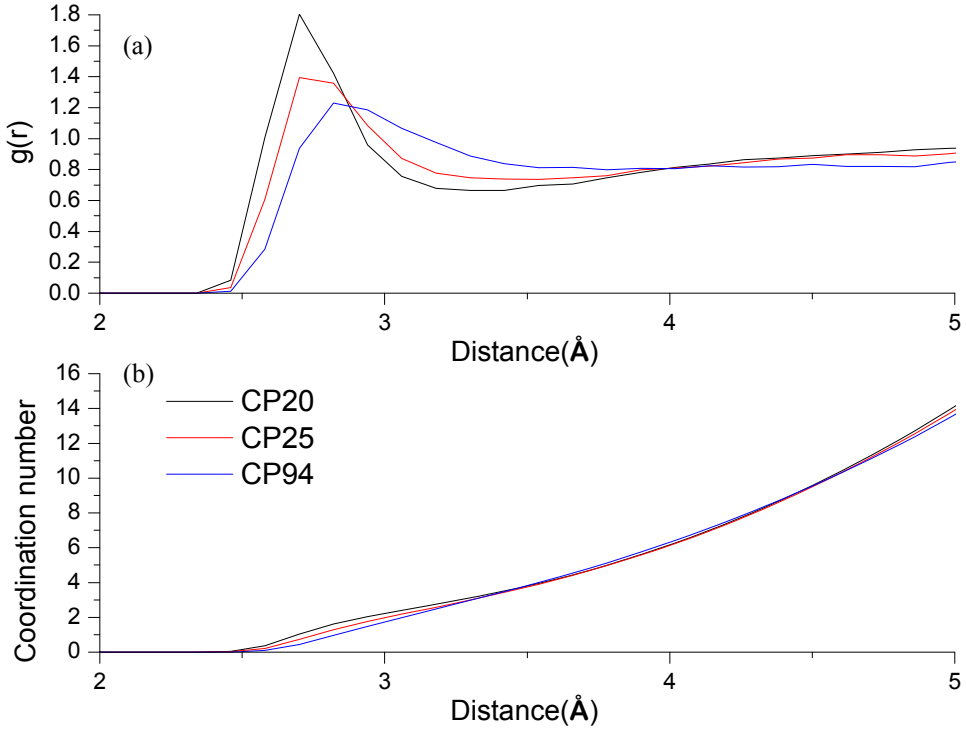


Figure 6-13 O_4–O (H_2O) radial distribution function (a) and coordination number (b) for the hydrated HPOs[a] during the whole 3 ns period.

[a] CP20 (black line), CP25 (red line), CP94 (blue line).

6.4 Discussion

Novel CHARMM force fields specific for the three HPOs were developed and shown to reproduce QM geometries, water interactions, vibrational spectra and PESs. The calculated MM dipole moments, however, are close to (not higher than) QM values probably because slightly different MM geometries were obtained. If the HPOs are surrounded by explicit H_2O molecules shielding the potential hydrogen bonding between O_3-H and O_4, the angle differences will be reduced. Although some of MM PESs have higher energy barriers (by greater than 1 kcal mol^{-1}) than QM ones, the positions of the maxima and minima are close, indicating that the MM geometries will be close to QM ones but with the corresponding MM dihedrals more rotationally constrained.

The optimisation of atom charges in CHARMM requires structures interacting with water molecules to reproduce their water interaction energies and distances from QM calculations. This methodology is superior to the fixed atom charges only determined from QM-computed electrostatic potentials in vacuum, if developed force fields are applied in MD simulations with explicit water molecules.

Comparison of the initially assigned parameters from the CHARMM website with the final optimised parameters (Tables 6-5, 6-6 for CP20, 6-11, 6-12 for CP25 and 6-17, 6-18 for CP94) demonstrates that the corresponding atom charges and the force constants of bonds, angles, dihedrals and improper dihedrals are significantly different from one another. The greatest difference in partial charge is seen for the 1-nitrogen atom on the HPO ring. The different partial charges indicate that there are no structures which are similar enough to HPOs within the current CHARMM force field database. HPOs possess the unique 3,4-oxygen atoms forming resonance structures with the 1-nitrogen atom and this makes the compounds significantly different from the catechols and pyridins in the CHARMM database. Because the unique resonance occurs within the HPO ring, the electron density at the 1-nitrogen atom is much less negative than the value predicted by the CHARMM website.

Comparison of the optimised parameters for the three different HPOs demonstrates that the atom charges are significantly different from each other, but for the same types of bonds, angles, dihedrals and improper dihedrals, the force constants are closely similar. This indicates

that methyl or alkyl groups can influence the electron density significantly within the structure but have a lesser effect on the compound's geometry.

The order of hydrogen bonding strength with H_2O for the three HPOs justified by RDFs follows the same order of HPO-water interaction energies (Tables 6-2, 6-8 and 6-14). The RDFs computed from the three successive 1 ns periods are not significantly different from each other, indicating that a 1 ns simulation is sufficient to model the HPO-water system. In an attempt to improve the statistics, all 3 ns simulations were averaged together to provide the final RDFs. The H_2O coordination number with an origin at O_3 for CP94 is found to be significantly lower than those calculated for the other two HPOs (Figure 6-12). This can be explained by CP94 possessing a 2-ethyl rather than the 2-methyl substituent present in the other two HPOs, the larger alkyl chain repelling significantly more H_2O molecules.

Chapter 7 Summary and Conclusions

The purpose of this project was to predict the physicochemical properties of 3-hydroxypyridin-4-one (HPO) Fe^{3+}-chelating agents using novel computational methods. The hydroxyl pK_a values of the HPOs were predicted using quantum mechanical (QM) static calculations with an experimentally fitted model to obtain results in excellent agreement with experimental data (with the mean of the absolute deviations from measured values ($|M|$) being ~0.20, compared to the estimated experimental errors of ~0.10). The QM static computation assists with the assignment of different pK_a values for the two oxygen atoms of HPOs (associated with experimental metal affinity determination), and this is particularly useful when the HPOs possess more than two pK_a values. The hydroxyl pK_a values of three alkyl HPOs were further predicted using QM molecular dynamic (MD) simulations to provide reasonable predictions ($|M|$ = 0.34) and additional insights into the atomistic details of the whole deprotonation process, and hydration configurations and dynamics. It was found that the distribution of substituted methyl groups around the HPO aromatic ring would appear to affect the surrounding water networking significantly and this impacts upon their pK_a values. The hydroxyl pK_a values were also predicted using existing QSPR commercial software and thermodynamic cycle calculations but yielded results of insufficient accuracy. In general, choice of pK_a prediction methods should depend on the purpose and the quantity of existing experimental data. For example, if more than a hundred compounds, with known pK_a values for a sufficient number of analogues, need screening for drug discovery, the QSPR methods are relatively cost effective. However, if compounds contain functional groups that possess hydrogen bonding donors and/or acceptors, or that exhibit strong electronic effect, the higher level computational methods are required. In the future, if the power of computers is improved significantly, QM-based methods can take much less computing time, thereby competing for the screening of a large number of compounds with high accuracy.

The (Fe^{3+}) log K_1 values of the HPOs were predicted in excellent agreement with experimental data ($|M|$ ~0.32, compared to the estimated experimental errors of ~0.15) using the analogous methodology as the QM static calculations with a regression model for the hydroxyl pK_a prediction. Prior to the construction of the prediction model, an expected linear relationship between the pK_as and (Fe^{3+}) log K_1 was utilised to verify the experimental (Fe^{3+}) log K_1 values. From this, ten previously reported HPOs with substituents carrying hydrogen bonding

donors/acceptors were found to possess incorrect (Fe^{3+}) log K_1 values and then were re-characterised using a 10 : 1 molar ratio of ligand : Fe^{3+}. To date, the QM method for predicting the metal affinity of chelators in absolute value has been only reported in the work detailed here. The methodology can in principle be applied to predict HPO (Fe^{3+}) log β_2 and log β_3 and this will require more computing power and time. In addition, special attention is required for the prediction of HPO (Fe^{3+}) log β_3 because the extra stabilising effect from solvent (entropy contributions) plays a significant role for the $[Fe^{3+}(HPO)_3]$ complexes or the corresponding hexadentate chelators. Thus, QM MD with explicit solvent molecules is suggested to investigate their hydration configurations and this requires extensive computing power.

With a view to the prediction of HPO membrane permeability, novel CHARMM force fields specific for three HPOs were parameterised and tested with explicit water molecules using classical MD simulations. Because of a limited research time, the actual membrane permeability values were not predicted. In the future, this can be carried out using constraints on the depth of HPO insertion in explicitly hydrated biological membranes through classical MD simulations, in which the constraint forces can be converted to free energies, and thus to membrane permeability values although extensive computing power is required.

In conclusion, there are many possibilities for carrying out a diversity of research with the increasing power of computers. Computational chemistry is a great tool for researchers to understand the molecular details of natural phenomena, which are often difficult to obtain through physical experiments. From this, humans can understand the world in depth and improve their quality of life.

References

- ACD/pKa DB 12.0, http://www.acdlabs.com/ (assessed February, 2011).
- CPMD, http://www.cpmd.org/. Copyright IBM Corp 1990-2008, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.
- HyperChem(TM) Professional 8.0, http://www.hyper.com/ (assessed February, 2011). 1115 NW 4th Street, Gainesville, Florida 32601, USA: Hypercube, Inc.
- Marvin 5.4.0.1, http://www.chemaxon.com/ (assessed February, 2011).
- Abeysinghe, R. D., Roberts, P. J., Cooper, C. E., MacLean, K. H., Hider, R. C. & Porter, J. B. (1996). The environment of the lipoxygenase iron binding site explored with novel hydroxypyridinone iron chelators. *J. Biol. Chem.* 271(14): 7965-7972.
- Albert, A. & Rubro, S. D. (1947). The influence of chemical constitution of antibacterial activity; a study of 8-hydroxyquinolin (oxine) and related compounds. *Br. J. Exp. Pathol.* 28(2): 69-87.
- Alderighi, L., Gans, P., Ienco, A., Peters, D., Sabatini, A. & Vacca, A. (1999). Hyperquad simulation and speciation (HySS): a utility program for the investigation of equilibria involving soluble and partially soluble species. *Coord. Chem. Rev.* 184: 311-318.
- Banin, E., Brady, K. M. & Greenberg, E. P. (2006). Chelator-induced dispersal and killing of Pseudomonas aeruginosa cells in a biofilm. *Appl. Environ. Microbiol.* 72(3): 2064-2069.
- Barton, J. C. & Edwards, C. Q. (2000). *Hemochromatosis : genetics, pathophysiology, diagnosis and treatment.* Cambridge: Cambridge University Press.
- Bergan, T., Klaveness, J. & Aasen, A. J. (2001). Chelating agents. *Chemotherapy* 47(1): 10-14.
- Chaberek, S. & Martell, A. E. (1959). *Organic sequestering agents : a discussion of the chemical behavior and applications of metal chelate compounds in aqueous systems.* Wiley; Chapman & Hall.
- Chandler, D. (1987). *Introduction to modern statistical mechanics.* Oxford University Press.

- Chen, Y. L., Barlow, D. J., Kong, X. L., Ma, Y. M. & Hider, R. C. (2012a). Prediction of 3-hydroxypyridin-4-one (HPO) hydroxyl pKa values. *Dalton Trans.* 41(21): 6549-6557.
- Chen, Y. L., Barlow, D. J., Kong, X. L., Ma, Y. M. & Hider, R. C. (2012b). Prediction of 3-hydroxypyridin-4-one (HPO) log K(1) values for Fe(iii). *Dalton Trans.* 41(35): 10784-10791.
- Chen, Y. L., Doltsinis, N. L., Hider, R. C. & Barlow, D. J. (2012c). Prediction of Absolute Hydroxyl pKa Values for 3-Hydroxypyridin-4-ones. *J. Phys. Chem. Lett.*: 2980-2985.
- Crichton, R. R. & Ward, R. J. (2006). *Metal-based neurodegeneration : from molecular mechanisms to therapeutic strategies.* Chichester : Wiley.
- Crisponi, G. & Remelli, M. (2008). Iron chelating agents for the treatment of iron overload. *Coord. Chem. Rev.* 252(10-11): 1225-1240.
- Crumbliss, A. L. & Harrington, J. M. (2009). Iron Sequestration by Small Molecules: Thermodynamic and Kinetic Studies of Natural Siderophores and Synthetic Model Compounds. *Adv. Inorg. Chem., Vol 61: Metal Ion Controlled Reactivity* 61: 179-250.
- Davies, J. E., Doltsinis, N. L., Kirby, A. J., Roussev, C. D. & Sprik, M. (2002). Estimating pK(a) values for pentaoxyphosphoranes. *J. Am. Chem. Soc.* 124(23): 6594-6599.
- Dehkordi, L. S., Liu, Z. D. & Hider, R. C. (2008). Basic 3-hydroxypyridin-4-ones: potential antimalarial agents. *Eur. J. Med. Chem.* 43(5): 1035-1047.
- Dhungana, S., White, P. S. & Crumbliss, A. L. (2001). Crystal structure of ferrioxamine B: a comparative analysis and implications for molecular recognition. *J. Biol. Inorg. Chem.* 6(8): 810-818.
- Fernandez-Serra, M. V. & Artacho, E. (2004). Network equilibration and first-principles liquid water. *J. Chem. Phys.* 121(22): 11136-11144.
- Foresman, J. B. & Frisch, Æ. (1996). *Exploring chemistry with electronic structure methods.* Pittsburgh, PA: Gaussian.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.

A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., M. W. Gill, P., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C. & Pople, J. A. (2004).Gaussian 03, Revision E.01. Wallingford CT: Gaussian, Inc.

- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J., J. A. , Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J. & Fox, D. J. (2009).Gaussian 09, Revision A.02 SMP. Wallingford CT: Gaussian, Inc.

- Gaeta, A. & Hider, R. C. (2005). The crucial role of metal ions in neurodegeneration: the basis for a promising therapeutic strategy. *Br. J. Pharmacol.* 146(8): 1041-1059.

- Gans, P., Sabatini, A. & Vacca, A. (1999). Determination of equilibrium constants from spectrophometric data obtained from solutions of known pH: The program pHab. *Annali. Di. Chimica.* 89(1-2): 45-49.

- Habgood, M. D., Liu, Z. D., Dehkordi, L. S., Khodr, H. H., Abbott, J. & Hider, R. C. (1999). Investigation into the correlation between the structure of hydroxypyridinones and blood-brain barrier permeability. *Biochem. Pharmacol.* 57(11): 1305-1310.

- Harris, D. C. & Aisen, P. (1973). Facilitation of Fe(II) autoxidation by Fe(3) complexing agents. *Biochim. Biophys. Acta* 329(1): 156-158.

- Hershko, C., Konijn, A. M. & Link, G. (1998). Iron chelators for thalassaemia. *Br. J. Haematol.* 101(3): 399-406.

- Hider, R. C. (1995). Potential protection from toxicity by oral iron chelators. *Toxicol. Lett.* 82-83: 961-967.

- Hider, R. C. & Hall, A. D. (1991). Clinically useful chelators of tripositive elements. *Prog. Med. Chem.* 28: 41-173.

- Hider, R. C. & Kong, X. (2010). Chemistry and biology of siderophores. *Nat. Prod. Rep.* 27(5): 637-657.

- Hider, R. C., Mohdnor, A. R., Silver, J., Morrison, I. E. G. & Rees, L. V. C. (1981). Model Compounds for Microbial Iron-Transport Compounds .1. Solution Chemistry and Mossbauer Study of Iron(Ii) and Iron(Iii) Complexes from Phenolic and Catecholic Systems. *Dalton Trans.* (2): 609-622.

- Hider, R. C., Roy, S., Ma, Y. M., Le Kong, X. & Preston, J. (2011). The potential application of iron chelators for the treatment of neurodegenerative diseases. *Metallomics* 3(3): 239-249.

- Hider, R. C., Singh, S. & Porter, J. B. (1992). Iron Chelating-Agents with Clinical Potential. *Proc. Roy. Soc. Edin. B-Biol. Sci.* 99: 137-168.

- Ho, J. & Coote, M. (2010). A universal approach for continuum solvent pKa calculations: are we there yet? *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoret. Chim. Acta)* 125(1): 3-21.

- Ho, J. M. & Coote, M. L. (2009). pK(a) Calculation of Some Biologically Important Carbon Acids - An Assessment of Contemporary Theoretical Procedures. *J. Chem. Theory Comput.* 5(2): 295-306.
- Hoffman, R. (1991). *Hematology : basic principles and practice.* Churchill Livingstone.
- Jensen, F. (2007). *Introduction to computational chemistry.* Chichester: Wiley.
- Jones, J. E. (1924). On the determination of molecular fields - II From the equation of state of a gas. *Proc. Roy. Soc. Series a-Containing Papers of a Mathematical and Physical Character* 106(738): 463-477.
- Kohn, W. & Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* 140(4A): 1133-&.
- Kong, X. (2008).Spectrophotometric Determination of Stability Constants of Iron Chelators. In *Pharmacy Department*, Vol. PhD Thesis: King's College London, University of London.
- Kong, X. L., Zhou, T., Neubert, H., Liu, Z. D. & Hider, R. C. (2006). 3-hydroxy-2-(5-hydroxypentyl)-4H-chromen-4-one: A bidentate or tridentate iron(III) ligand? *J. Med. Chem.* 49(10): 3028-3031.
- Liao, C. & Nicklaus, M. C. (2009). Comparison of nine programs predicting pK(a) values of pharmaceutical substances. *J. Chem. Inf. Model* 49(12): 2801-2812.
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* 23(1-3): 3-25.
- Liu, Z. D. & Hider, R. C. (2002). Design of iron chelators with therapeutic application. *Coord. Chem. Rev.* 232(1-2): 151-171.
- Liu, Z. D., Kayyali, R., Hider, R. C., Porter, J. B. & Theobald, A. E. (2002). Design, synthesis, and evaluation of novel 2-substituted 3-hydroxypyridin-4-ones: structure-activity investigation of metalloenzyme inhibition by iron chelators. *J. Med. Chem.* 45(3): 631-639.
- Liu, Z. D., Khodr, H. H., Liu, D. Y., Lu, S. L. & Hider, R. C. (1999). Synthesis, physicochemical characterization, and biological evaluation of 2-(1'-hydroxyalkyl)-3-

hydroxypyridin-4-ones: novel iron chelators with enhanced pFe(3+) values. *J. Med. Chem.* 42(23): 4814-4823.

- Lowe, M. B. & Phillips, J. N. (1962). A Possible Mode of Action of Some Anti-Fungal and Anti-Bacterial Chelating Agents. *Nature* 194(4833): 1058-&.

- Ma, Y., Roy, S., Kong, X., Chen, Y., Liu, D. & Hider, R. C. (2012). Design and synthesis of fluorinated iron chelators for metabolic study and brain uptake. *J. Med. Chem.* 55(5): 2185-2195.

- Martell, A. E. & Smith, R. M. (1974-1989). *Critical Stability Constants.* London: Plenum Press.

- Merkofer, M., Kissner, R., Hider, R. C. & Koppenol, W. H. (2004). Redox properties of the iron complexes of orally active iron chelators CP20, CP502, CP509, and ICL670. *Helv. Chim. Acta* 87(12): 3021-3034.

- Pardridge, W. M. (2012). Drug transport across the blood-brain barrier. *J. Cereb. Blood Flow Metab.*

- Perdew, J. P., Burke, K. & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77(18): 3865-3868.

- Pietrangelo, A. (2007). Iron chelation beyond transfusion iron overload. *Am. J. Hematol.* 82(12 Suppl): 1142-1146.

- Pippard, M. J., Callender, S. T., Warner, G. T. & Weatherall, D. J. (1979). Iron absorption and loading in beta-thalassaemia intermedia. *Lancet* 2(8147): 819-821.

- Piyamongkol, S., Ma, Y. M., Kong, X. L., Liu, Z. D., Aytemir, M. D., van der Helm, D. & Hider, R. C. (2010). Amido-3-hydroxypyridin-4-ones as iron(III) ligands. *Chemistry* 16(21): 6374-6381.

- Porter, J. B. (1997). A risk-benefit assessment of iron-chelation therapy. *Drug Saf.* 17(6): 407-421.

- Rai, B. L., Dekhordi, L. S., Khodr, H., Jin, Y., Liu, Z. & Hider, R. C. (1998). Synthesis, physicochemical properties, and evaluation of N-substituted-2-alkyl-3-hydroxy-4(1H)-pyridinones. *J. Med. Chem.* 41(18): 3347-3359.

- Reed, W. & Vichinsky, E. P. (1998). New considerations in the treatment of sickle cell disease. *Annu. Rev. Med.* 49: 461-474.

- Roux, B. & Simonson, T. (1999). Implicit solvent models. *Biophys. Chem.* 78(1-2): 1-20.

- Schwegler, E., Grossman, J. C., Gygi, F. & Galli, G. (2004). Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II. *J. Chem. Phys.* 121(11): 5400-5409.

- Sigel, A., Sigel, H. & Sigel, R. K. O. (2006). *Neurodegenerative diseases and metal ions.* Chichester: Wiley.

- Sit, P. H. L. & Marzari, N. (2005). Static and dynamical properties of heavy water at ambient conditions from first-principles molecular dynamics. *J. Chem. Phys.* 122(20): 204510.

- Sprik, M. & Ciccotti, G. (1998). Free energy from constrained molecular dynamics. *J. Chem. Phys.* 109(18): 7737-7744.

- Valdebenito, M., Crumbliss, A. L., Winkelmann, G. & Hantke, K. (2006). Environmental factors influence the production of enterobactin, salmochelin, aerobactin, and yersiniabactin in Escherichia coli strain Nissle 1917. *Int. J. Med. Microbiol.* 296(8): 513-520.

- van Gunsteren, W. F. & Mark, A. E. (1998). Validation of molecular dynamics simulation. *J. Chem. Phys.* 108(15): 6109-6116.

- Vanderbilt, D. (1990). Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B* 41(11): 7892-7895.

- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & Mackerell, A. D., Jr. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31(4): 671-690.

- Weinreb, O., Amit, T., Mandel, S., Kupershmidt, L. & Youdim, M. B. (2010). Neuroprotective multifunctional iron chelators: from redox-sensitive process to novel therapeutic opportunities. *Antioxid Redox Signal* 13(6): 919-949.

- Xiao, G. Y., Vanderhelm, D., Hider, R. C. & Dobbin, P. S. (1992). Structure Stability Relationships of 3-Hydroxypyridin-4-One Complexes. *Dalton Trans.* (22): 3265-3271.

- Zhou, T., Ma, Y., Kong, X. & Hider, R. C. (2012). Design of iron chelators with therapeutic application. *Dalton Trans.* 41(21): 6371-6389.

Appendix A UV/Vis spectra and speciation plots

A.1 CP28

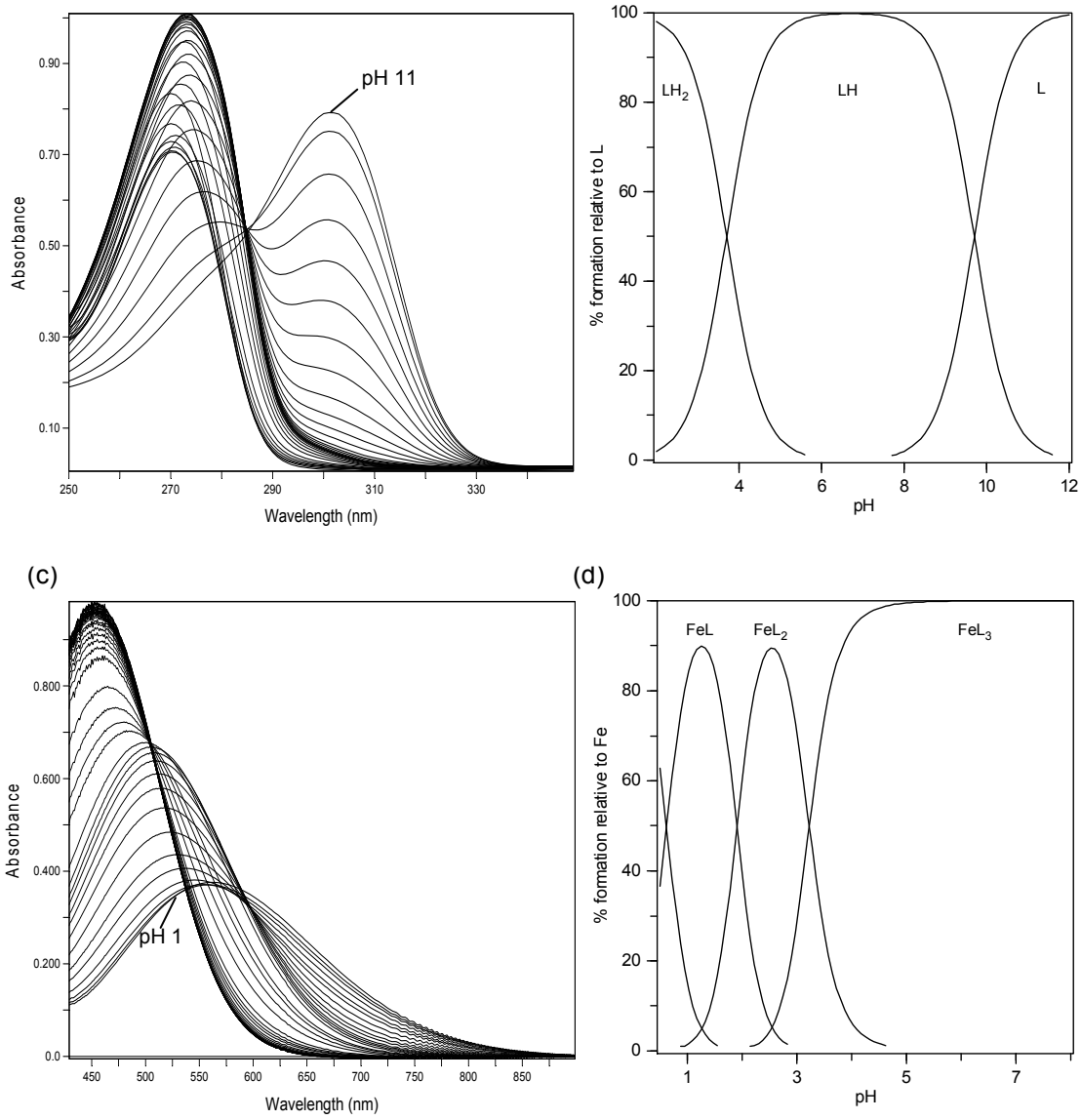


Figure A - 1 UV/Vis spectra and corresponding speciation plots for CP28.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1476.9 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 492.1 μM and [Fe^{3+}] = 46.5 μM over the pH range 1-8.

A.2 CP38

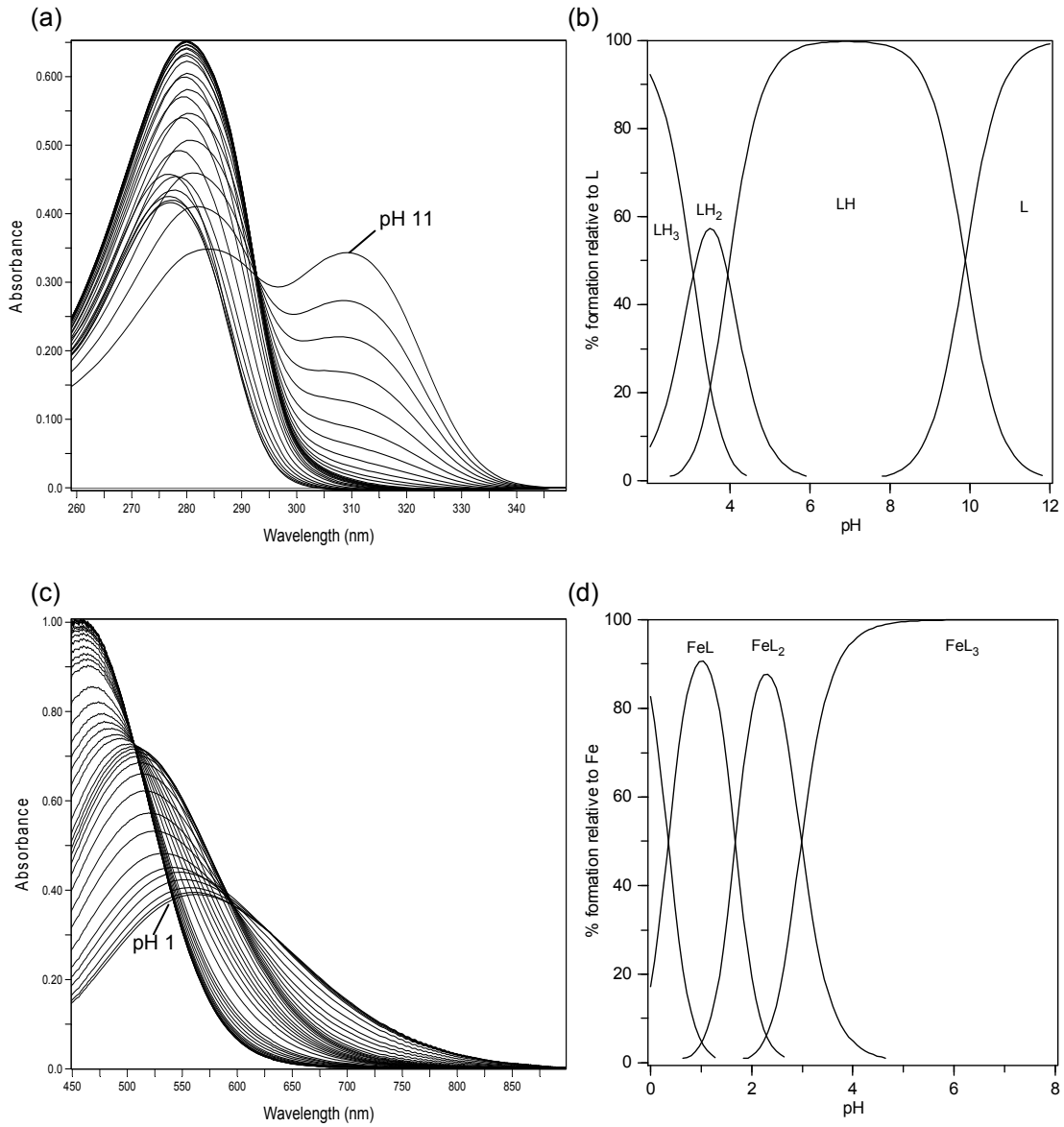


Figure A - 2 UV/Vis spectra and corresponding speciation plots for CP38.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 682.9 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 462.3 µM and [Fe^{3+}] = 45.7 µM over the pH range 1-8.

A.3 CP40

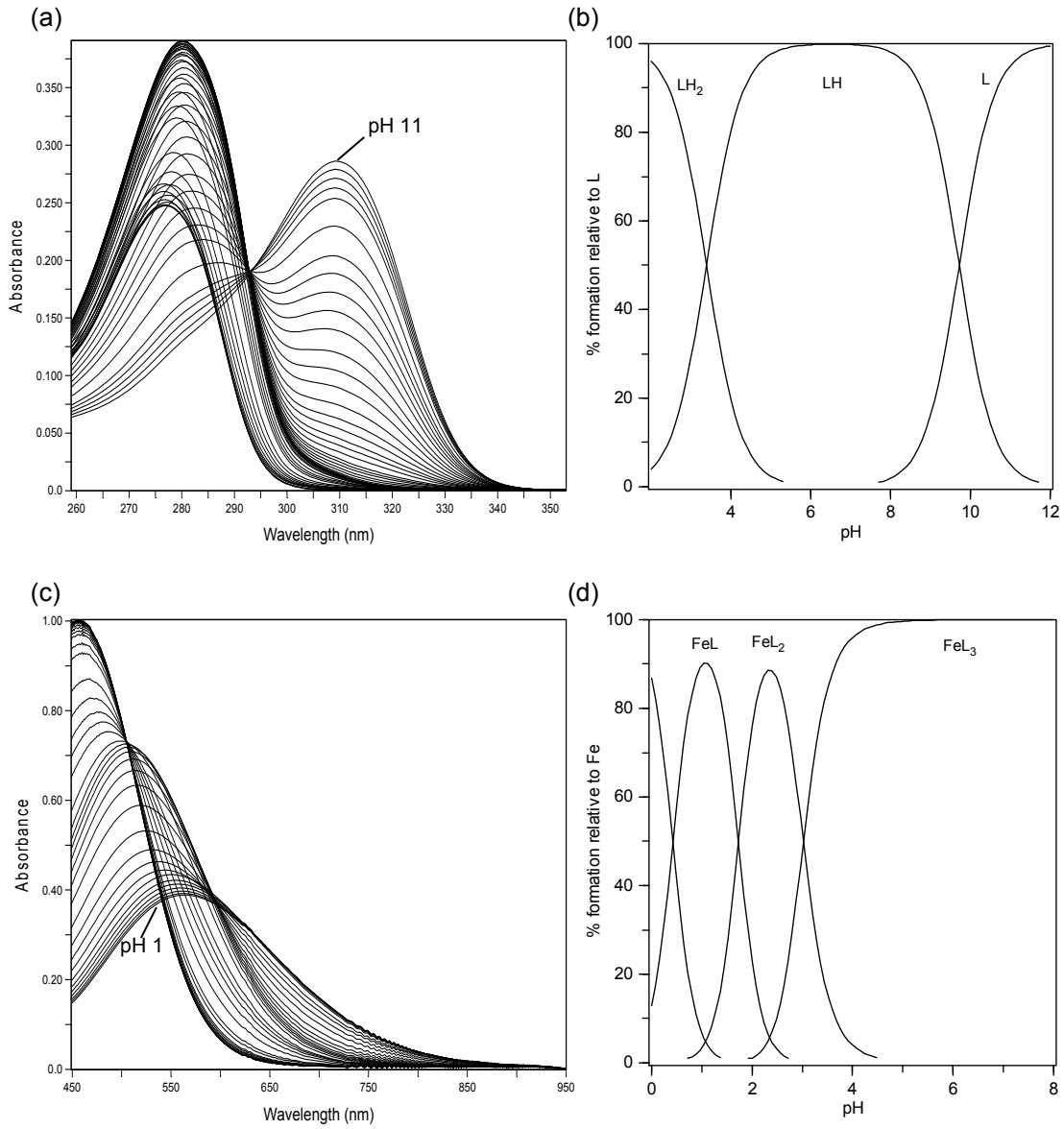


Figure A - 3 UV/Vis spectra and corresponding speciation plots for CP40.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 316.7 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 471.3 μM and [Fe^{3+}] = 46.0 μM over the pH range 1-8.

A.4 CP70

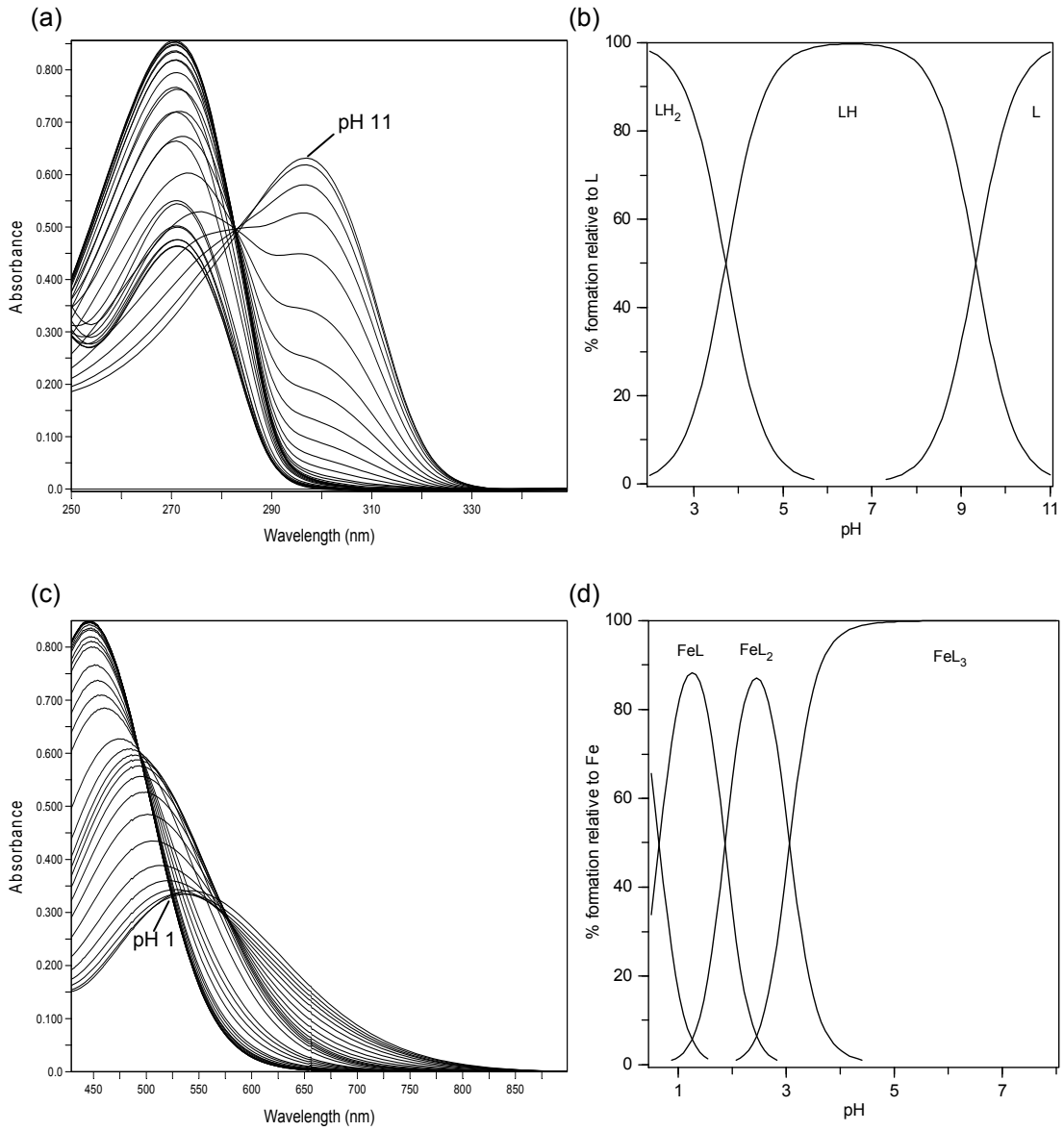


Figure A - 4 UV/Vis spectra and corresponding speciation plots for CP70.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 316.7 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 471.3 μM and [Fe^{3+}] = 46.0 μM over the pH range1-8 .

A.5 CP102

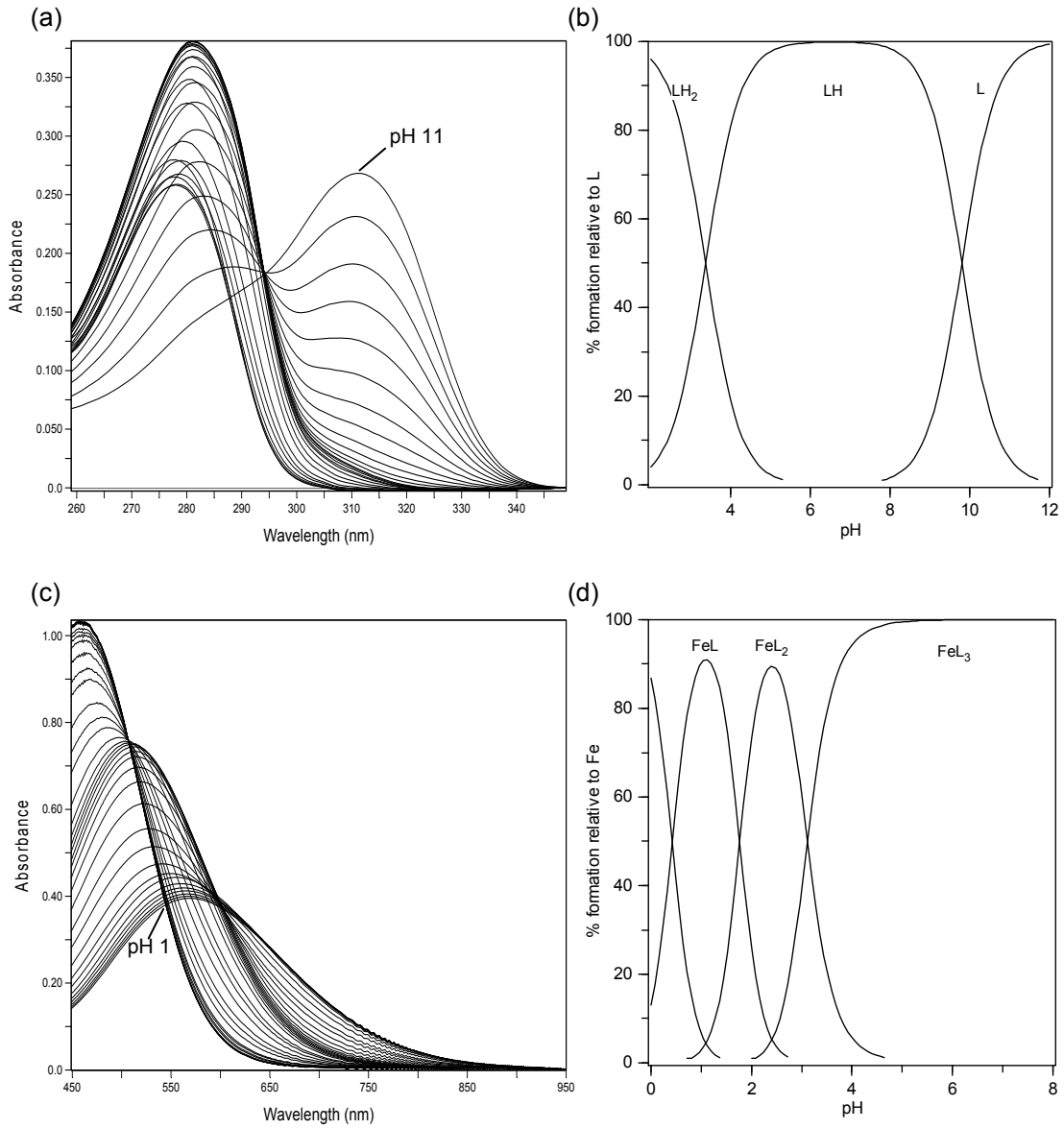


Figure A - 5 UV/Vis spectra and corresponding speciation plots for CP102.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 366.4 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 462.0 µM and [Fe^{3+}] = 45.9 µM over the pH range 1-8.

A.6 CP110

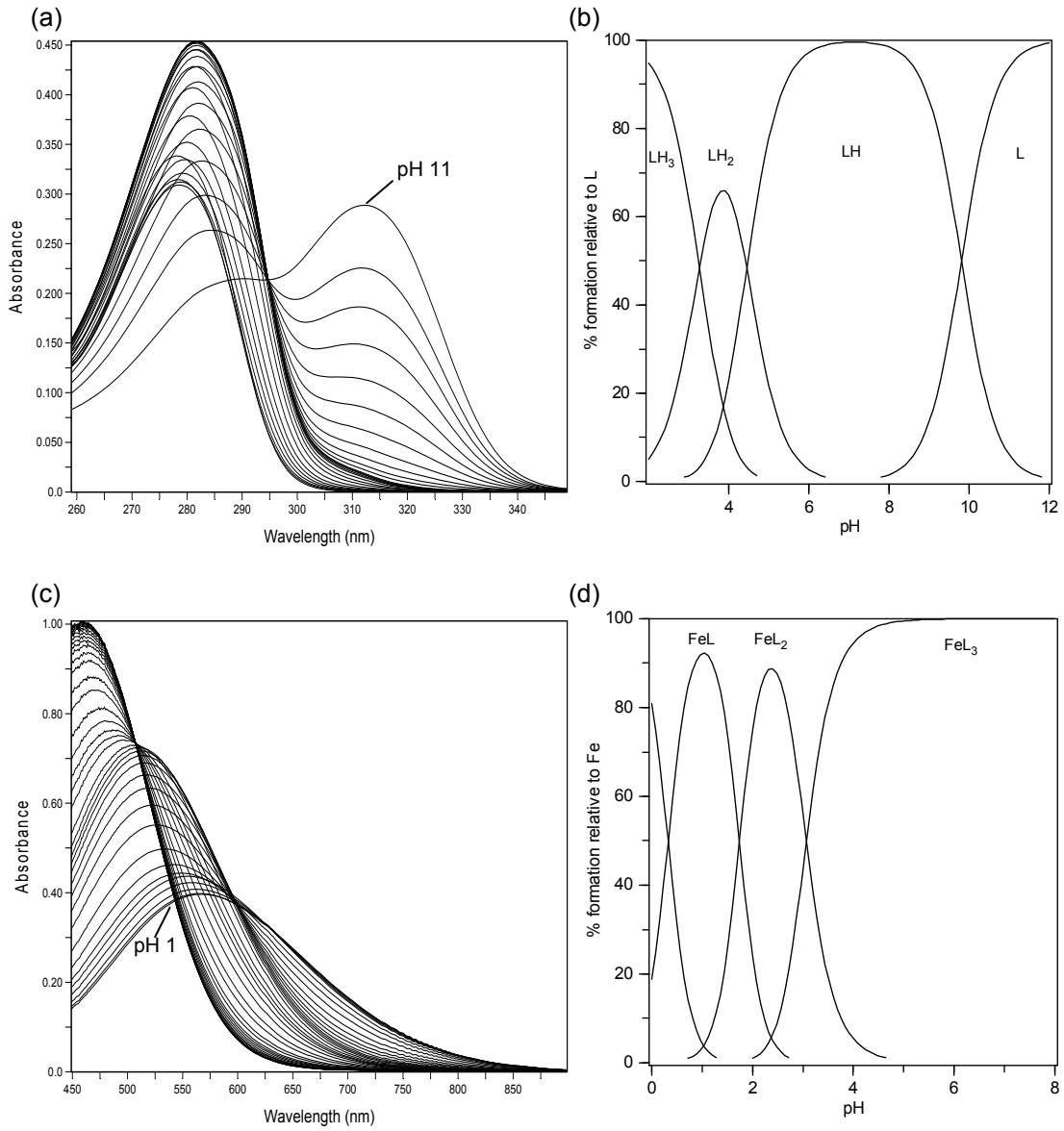


Figure A - 6 UV/Vis spectra and corresponding speciation plots for CP110.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 492.2 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 462.0 μM and [Fe^{3+}] = 45.9 μM over the pH range 1-8.

A.7 CP111

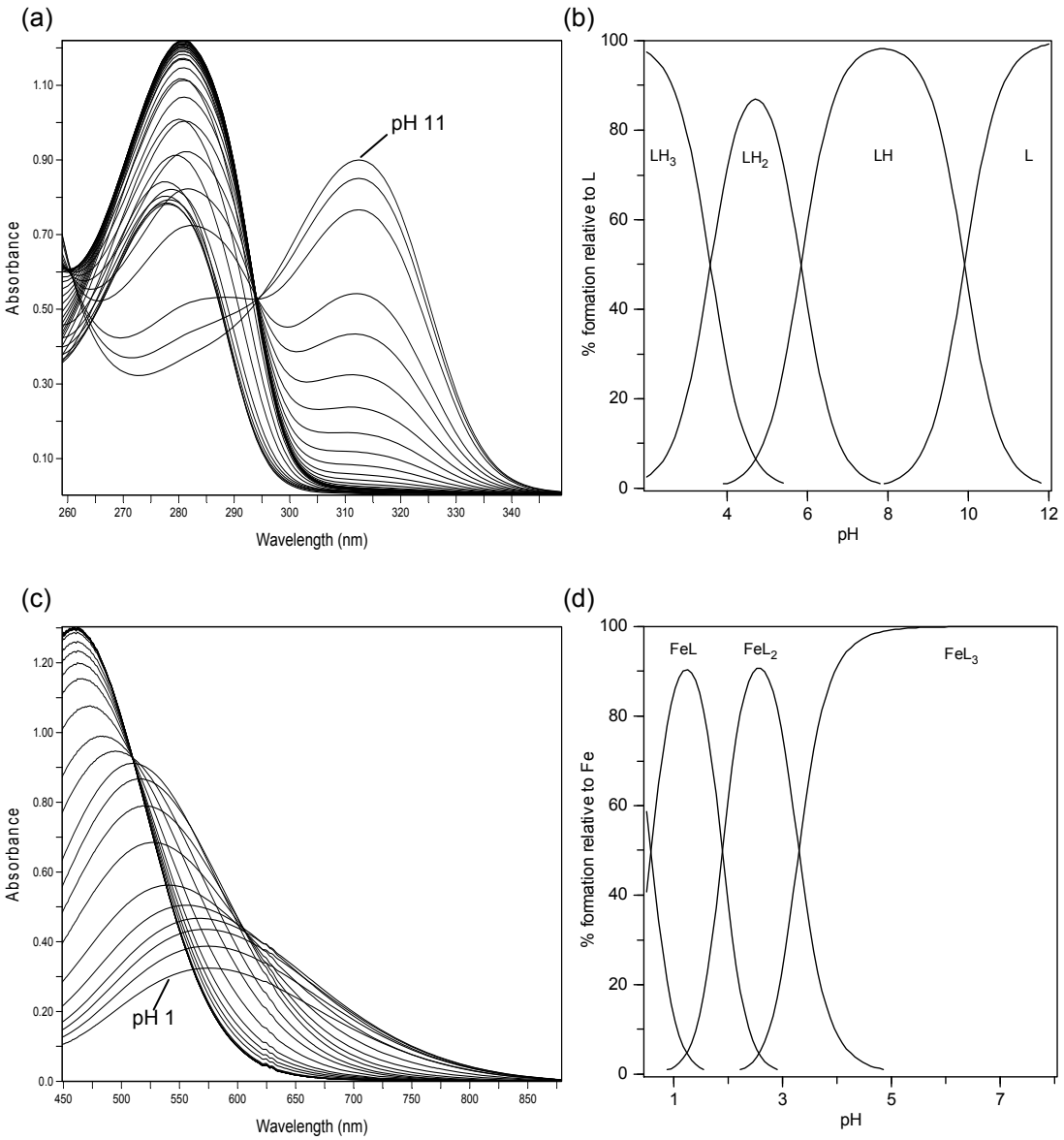


Figure A - 7 UV/Vis spectra and corresponding speciation plots for CP111.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 29.0 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 582.5 µM and [Fe^{3+}] = 56.0 µM over the pH range 1-8.

A.8 CP352

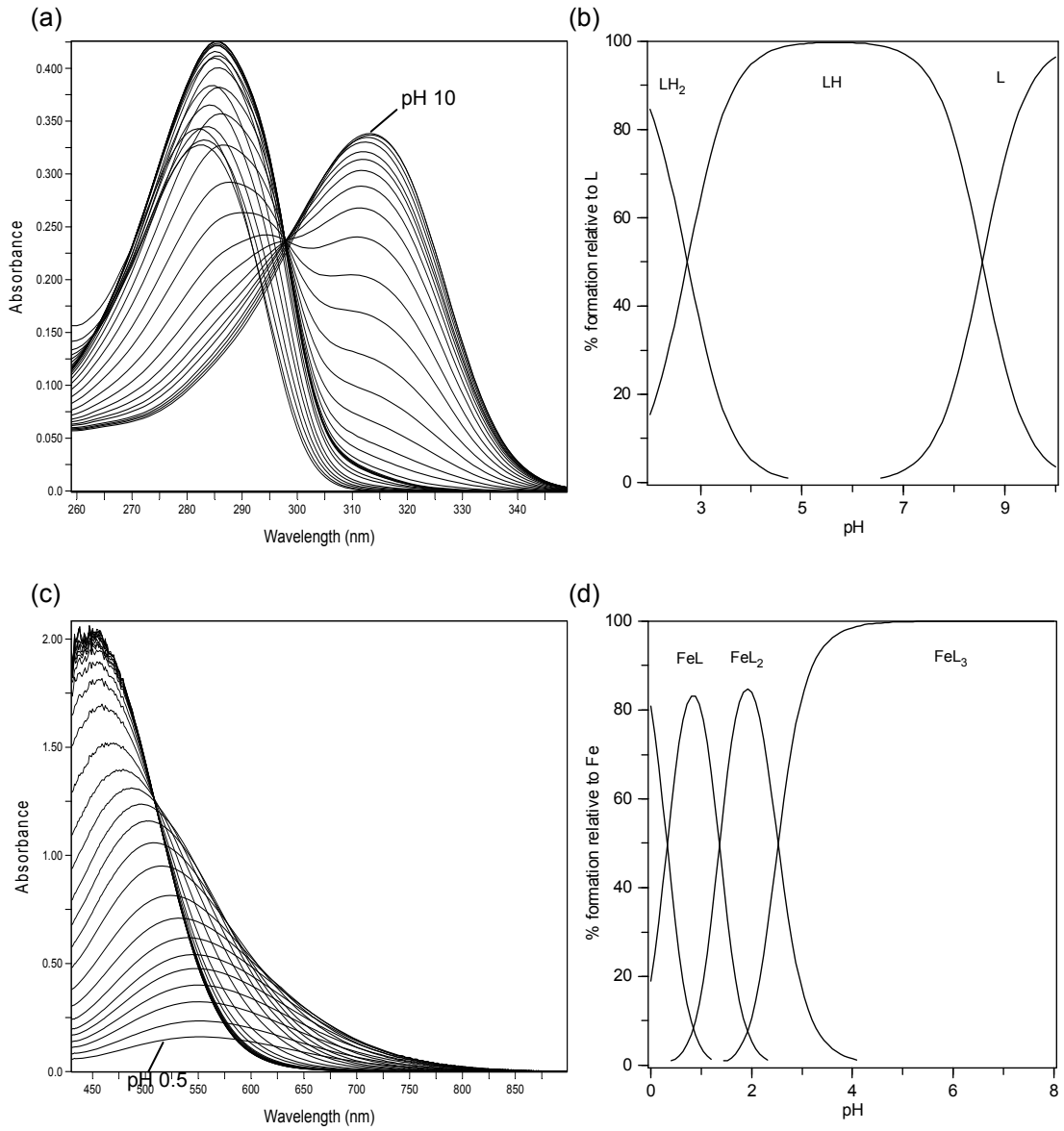


Figure A - 8 UV/Vis spectra and corresponding speciation plots for CP352.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 349.2 μM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 455.0 μM and [Fe^{3+}] = 45.1 μM over the pH range 0.5-8.

A.9 CP359

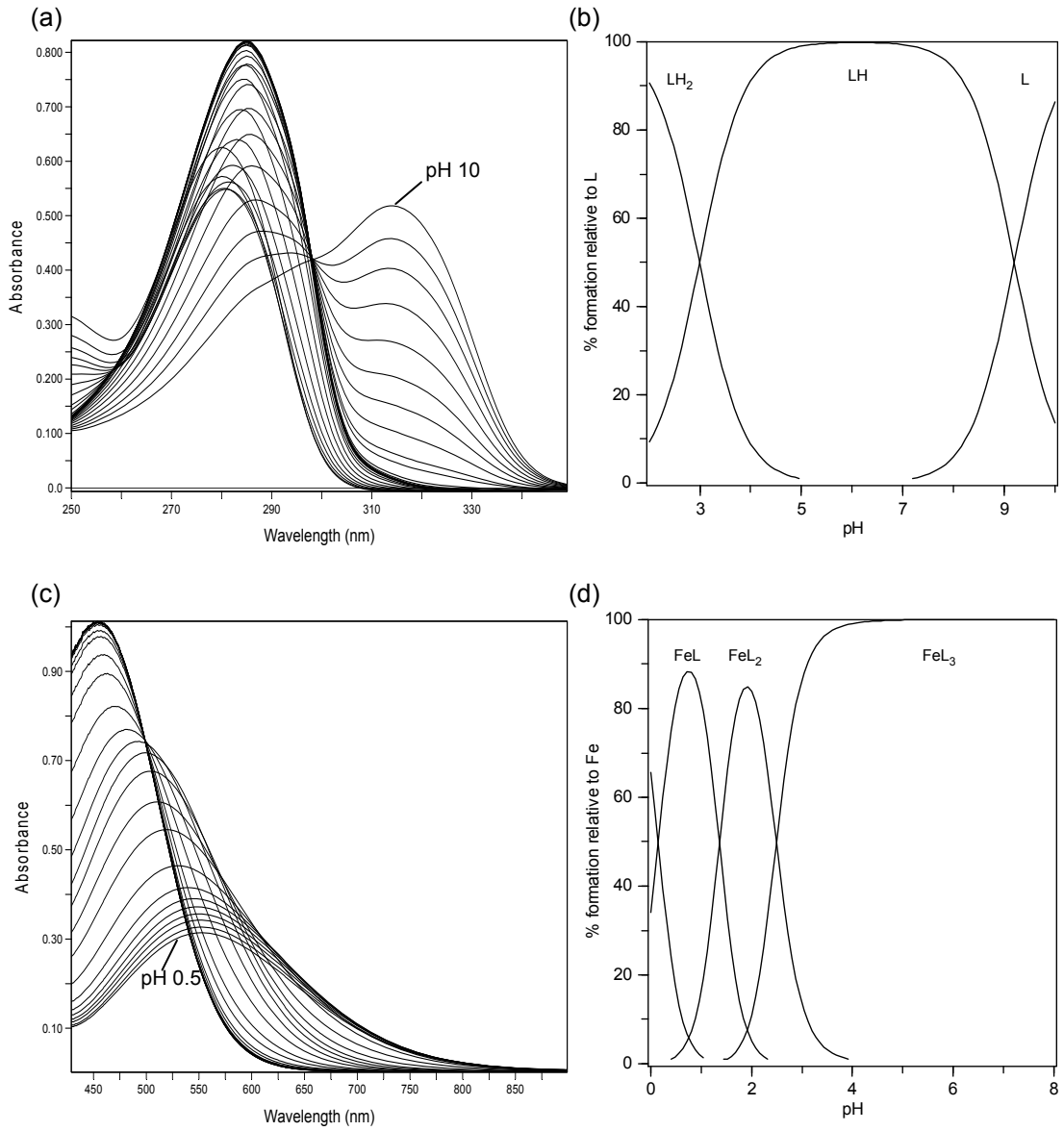


Figure A - 9 UV/Vis spectra and corresponding speciation plots for CP359.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 789.8 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 458.5 µM and [Fe^{3+}] = 45.9 µM over the pH range 0.5-8.

A.10 CP364

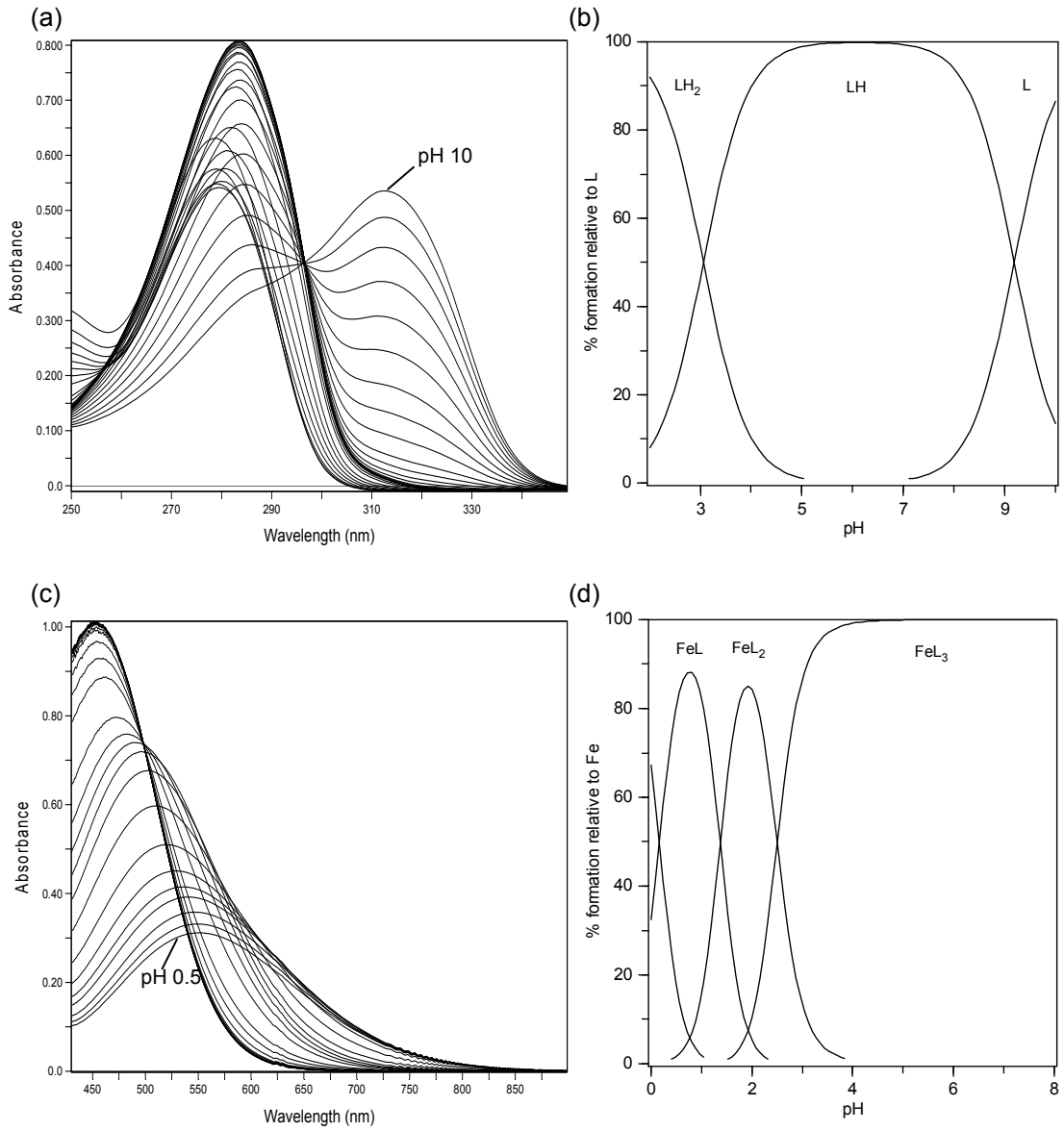


Figure A - 10 UV/Vis spectra and corresponding speciation plots for CP364.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 830.4 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 466.0 µM and [Fe^{3+}] = 45.9 µM over the pH range 0.5-8.

A.11 CP365

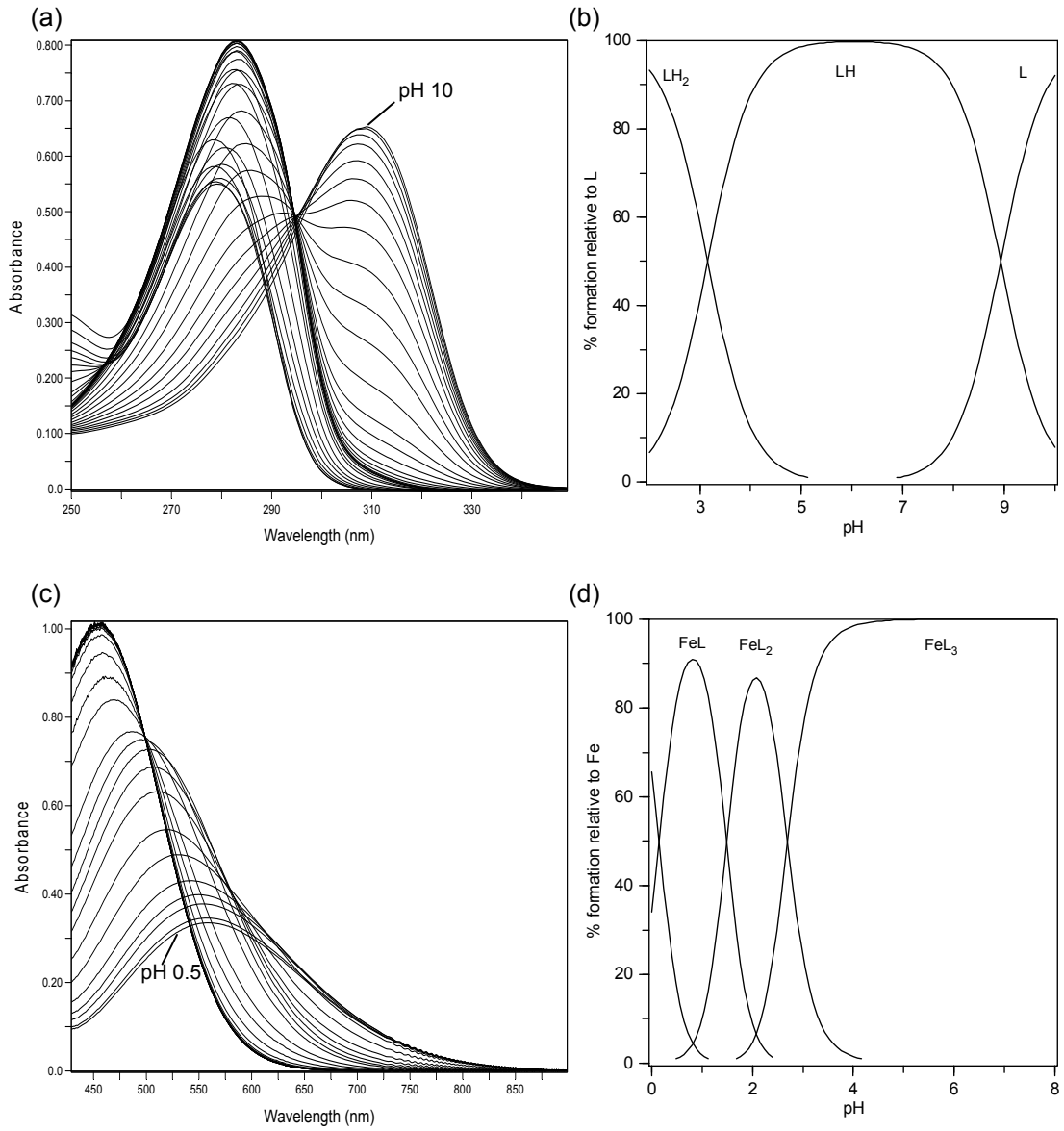


Figure A - 11 UV/Vis spectra and corresponding speciation plots for CP365.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 762.0 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 458.6 µM and [Fe^{3+}] = 46.1 µM over the pH range 0.5-8.

A.12 CP366

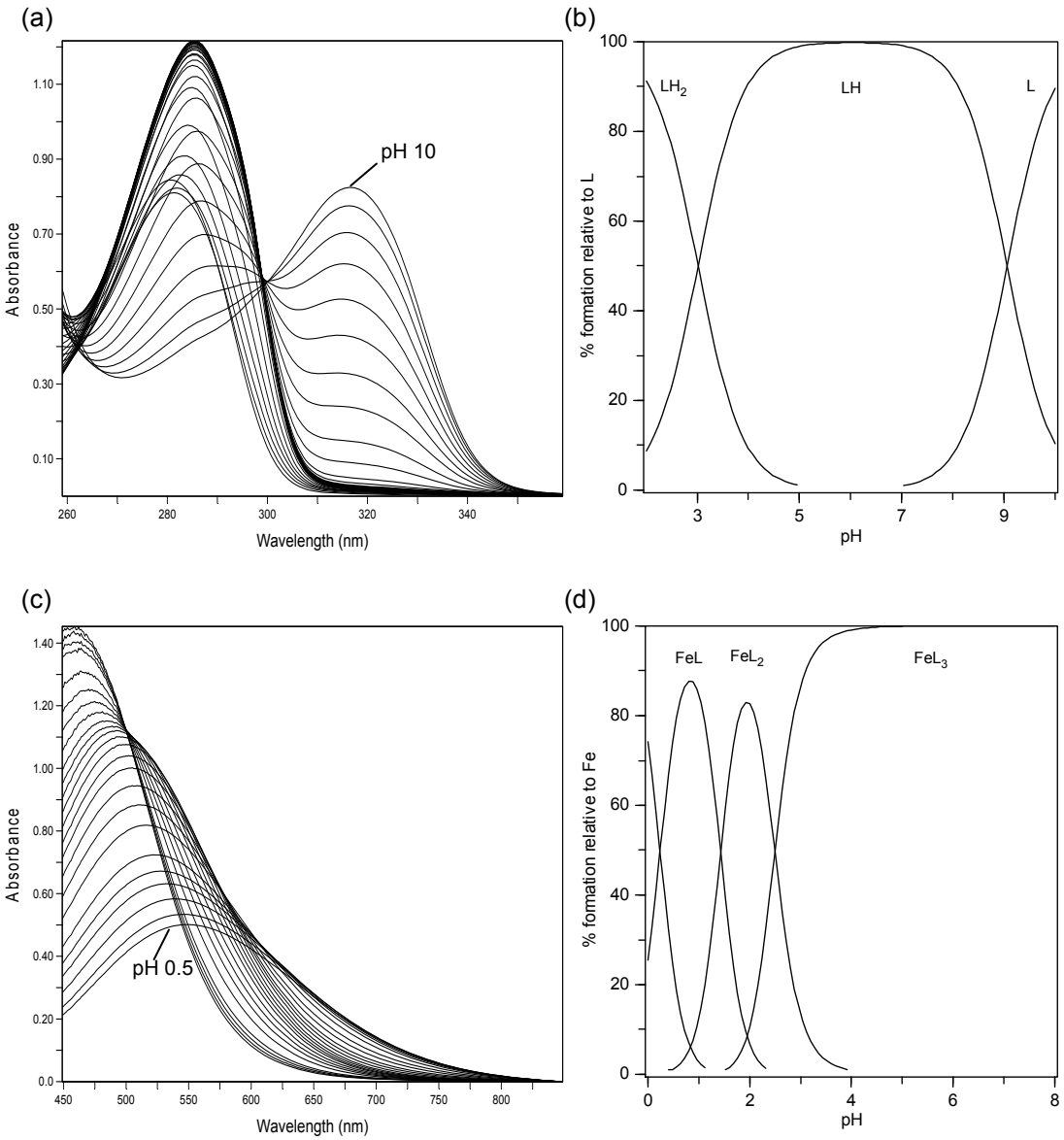


Figure A - 12 UV/Vis spectra and corresponding speciation plots for CP366.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 33.1 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 639.0 µM and [Fe^{3+}] = 61.4 µM over the pH range 0.5-8.

A.13 CP370

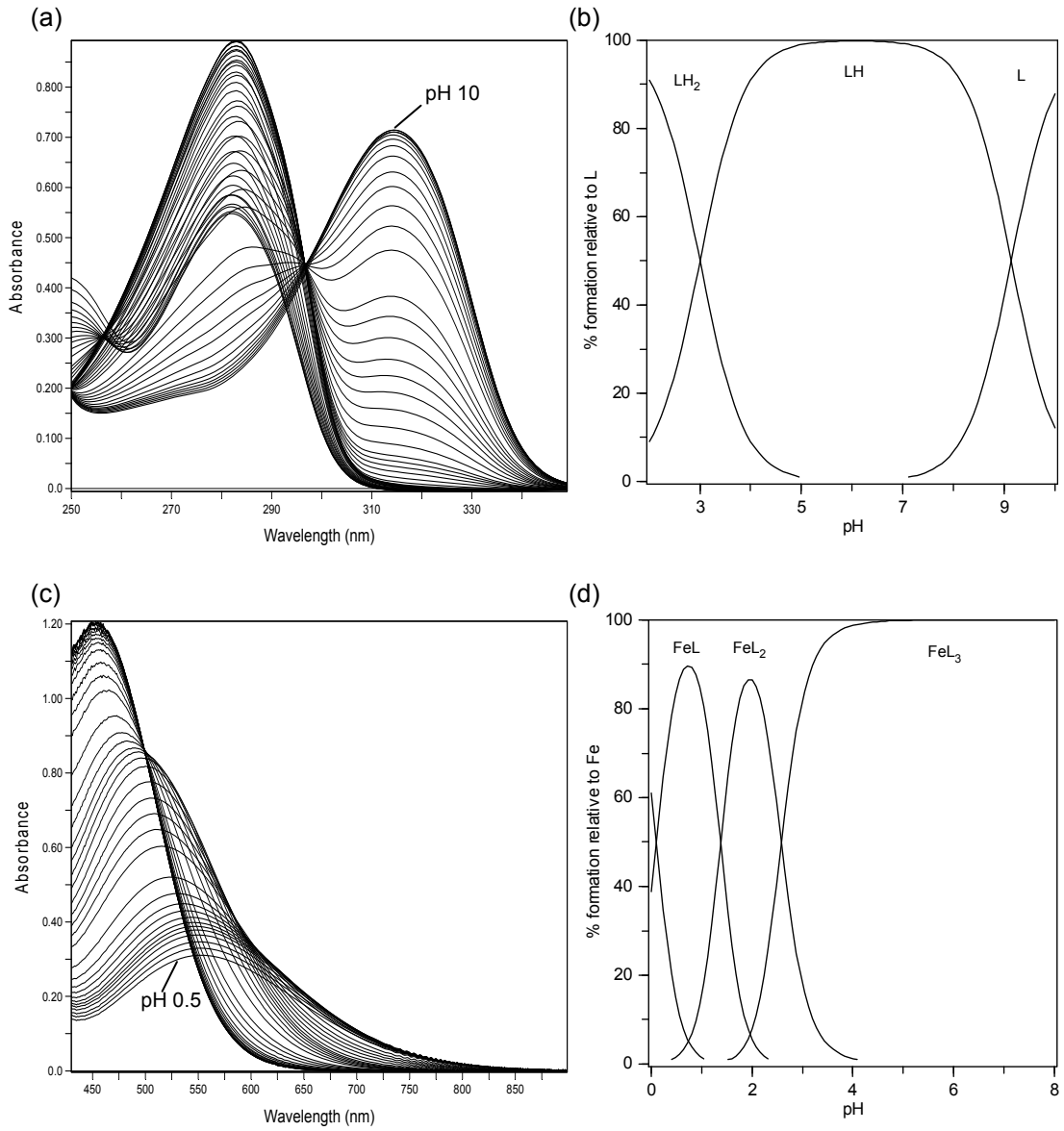


Figure A - 13 UV/Vis spectra and corresponding speciation plots for CP370.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 805.0 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 468.8 µM and [Fe^{3+}] = 46.1 µM over the pH range 0.5-8.

A.14 CP372

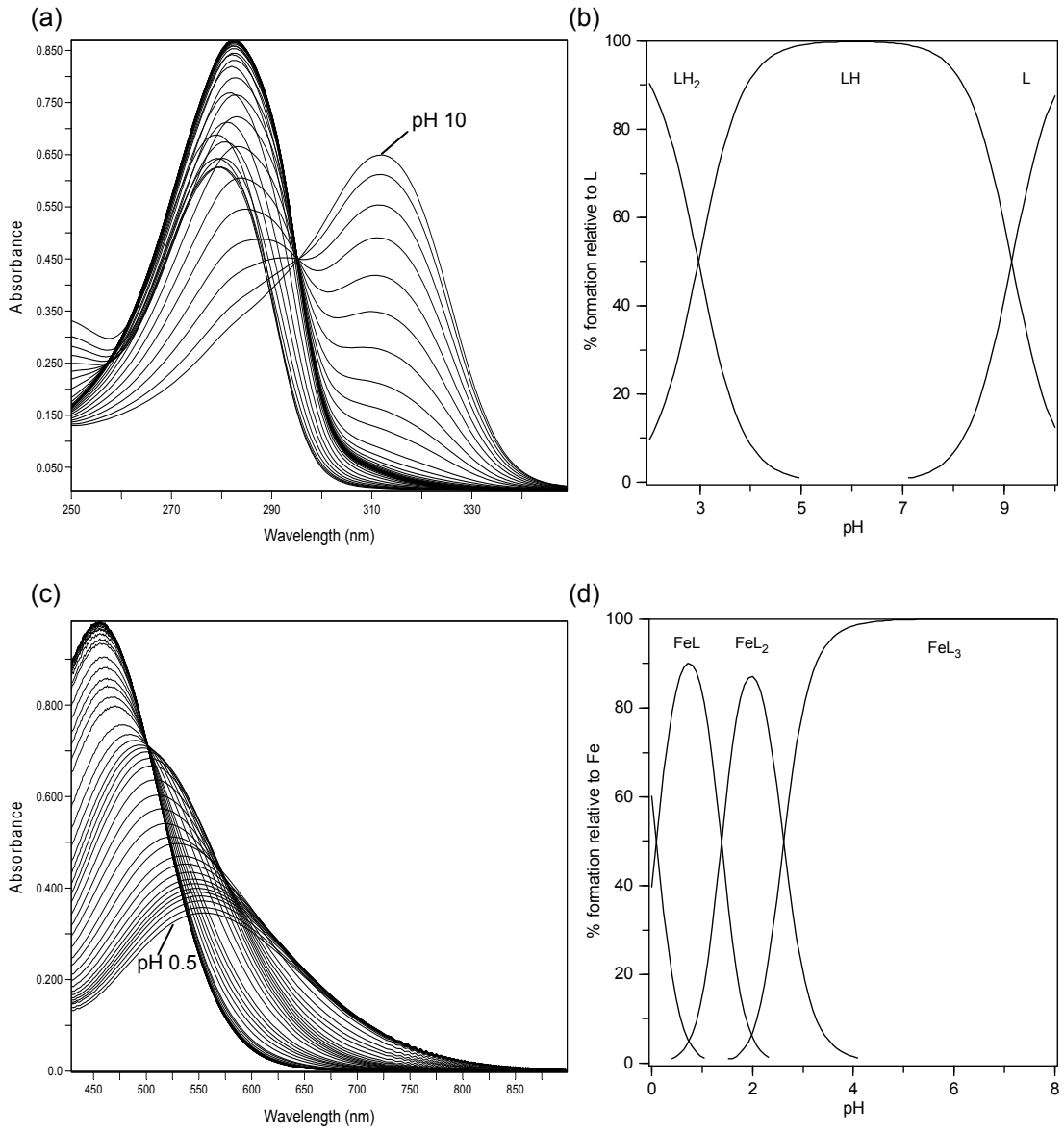


Figure A - 14 UV/Vis spectra and corresponding speciation plots for CP372.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 778.4 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 464.4 µM and [Fe^{3+}] = 45.8 µM over the pH range 0.5-8.

A.15 CP374

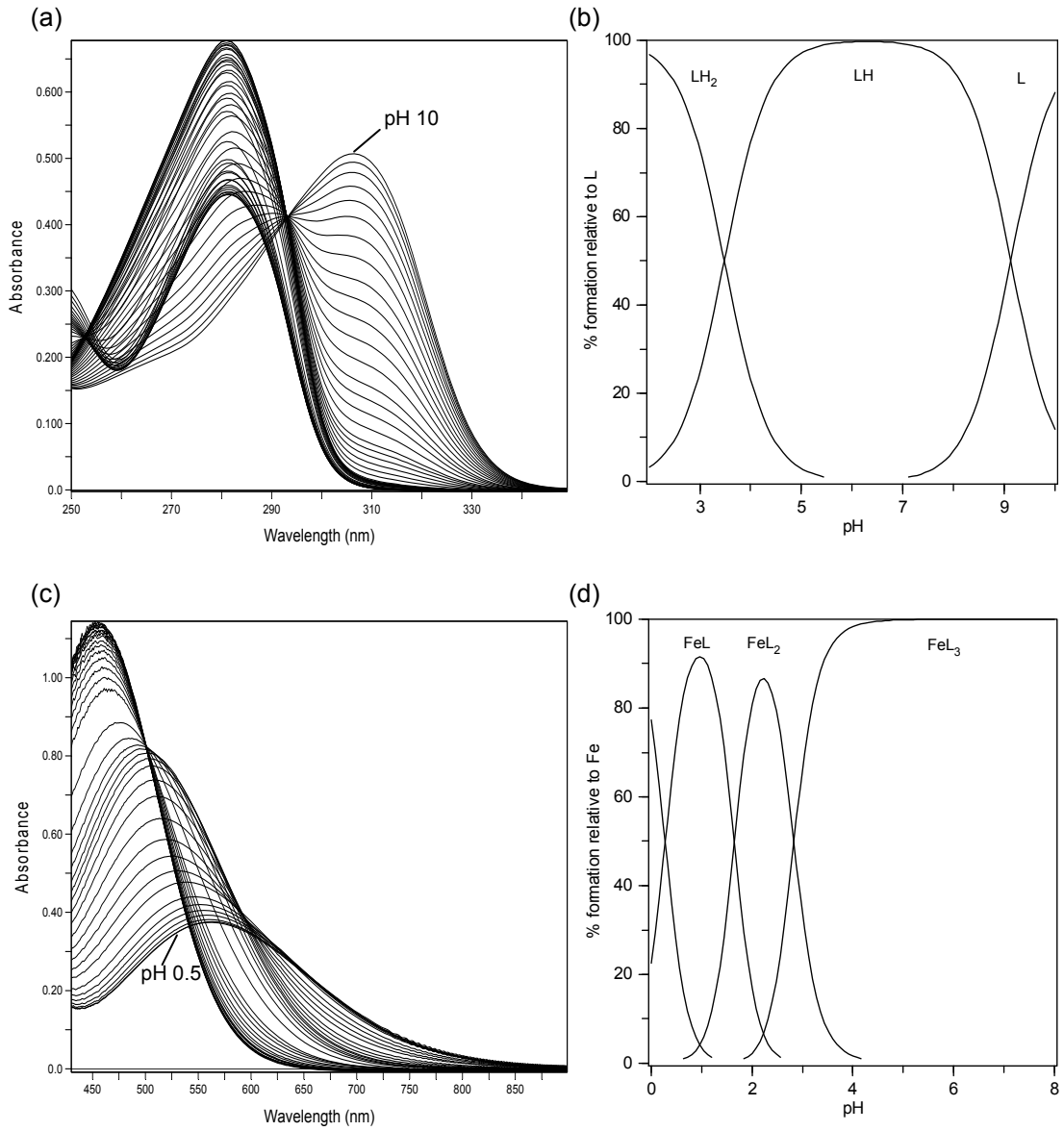


Figure A - 15 UV/Vis spectra and corresponding speciation plots for CP374.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 634.8 µM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 470.6 µM and [Fe^{3+}] = 46.3 µM over the pH range 0.5-8.

A.16 CP375

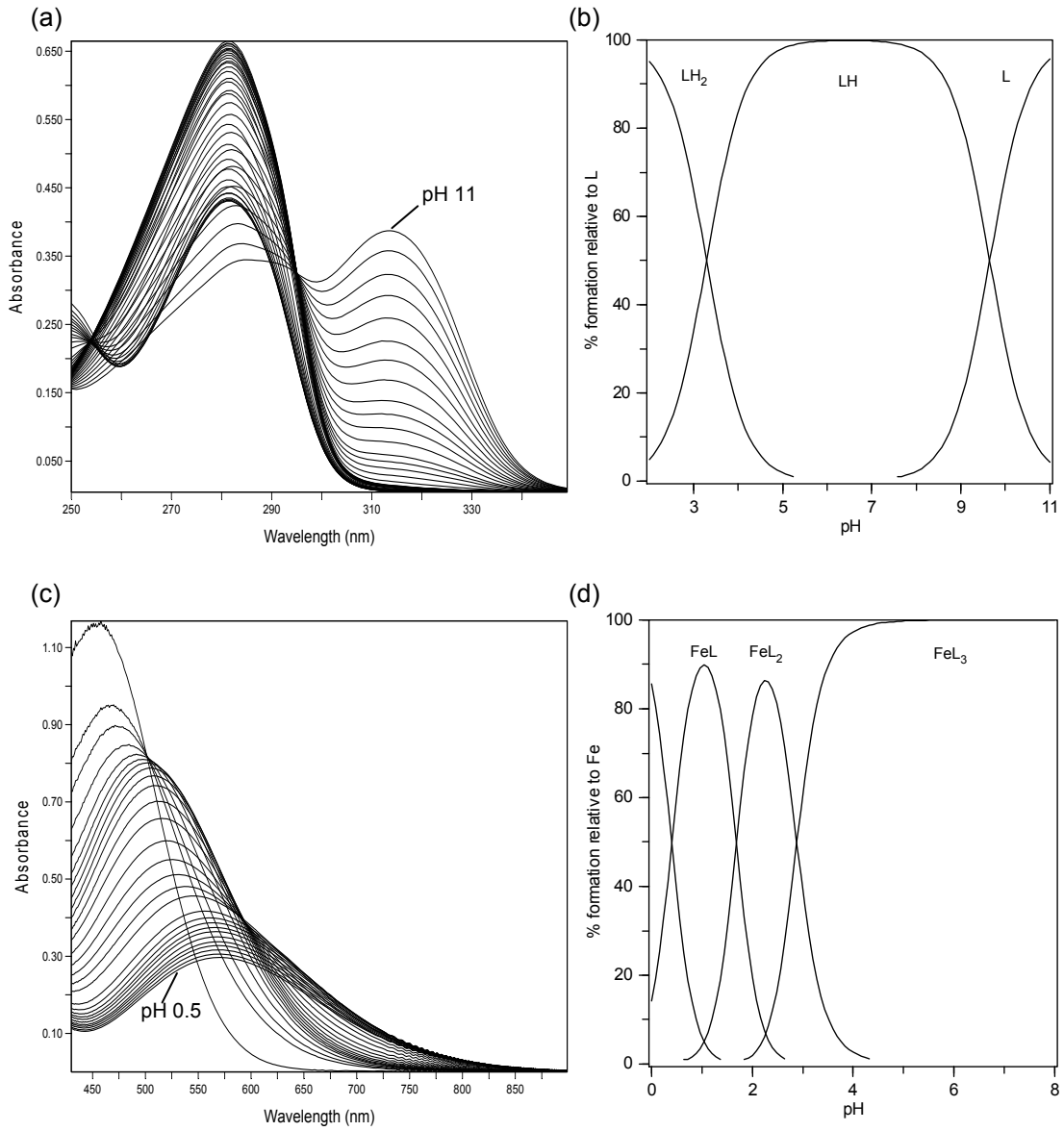


Figure A - 16 UV/Vis spectra and corresponding speciation plots for CP375.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 568.6 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 462.3 μM and [Fe^{3+}] = 46.0 μM over the pH range 0.5-8.

A.17 CP414

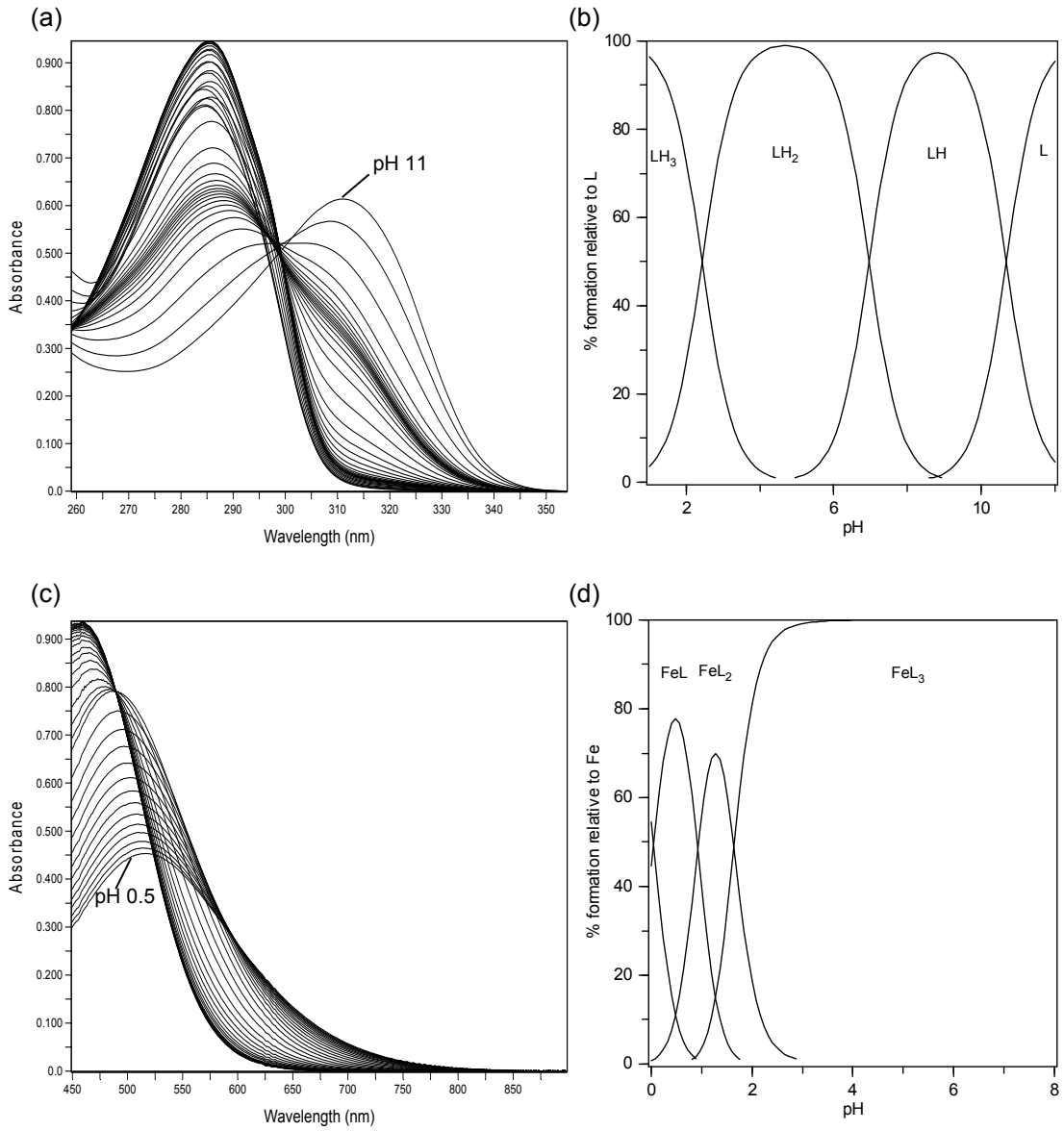


Figure A - 17 UV/Vis spectra and corresponding speciation plots for CP414.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 137.6 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 538.9 µM and [Fe^{3+}] = 44.7 µM over the pH range 0.5-8.

A.18 CP510

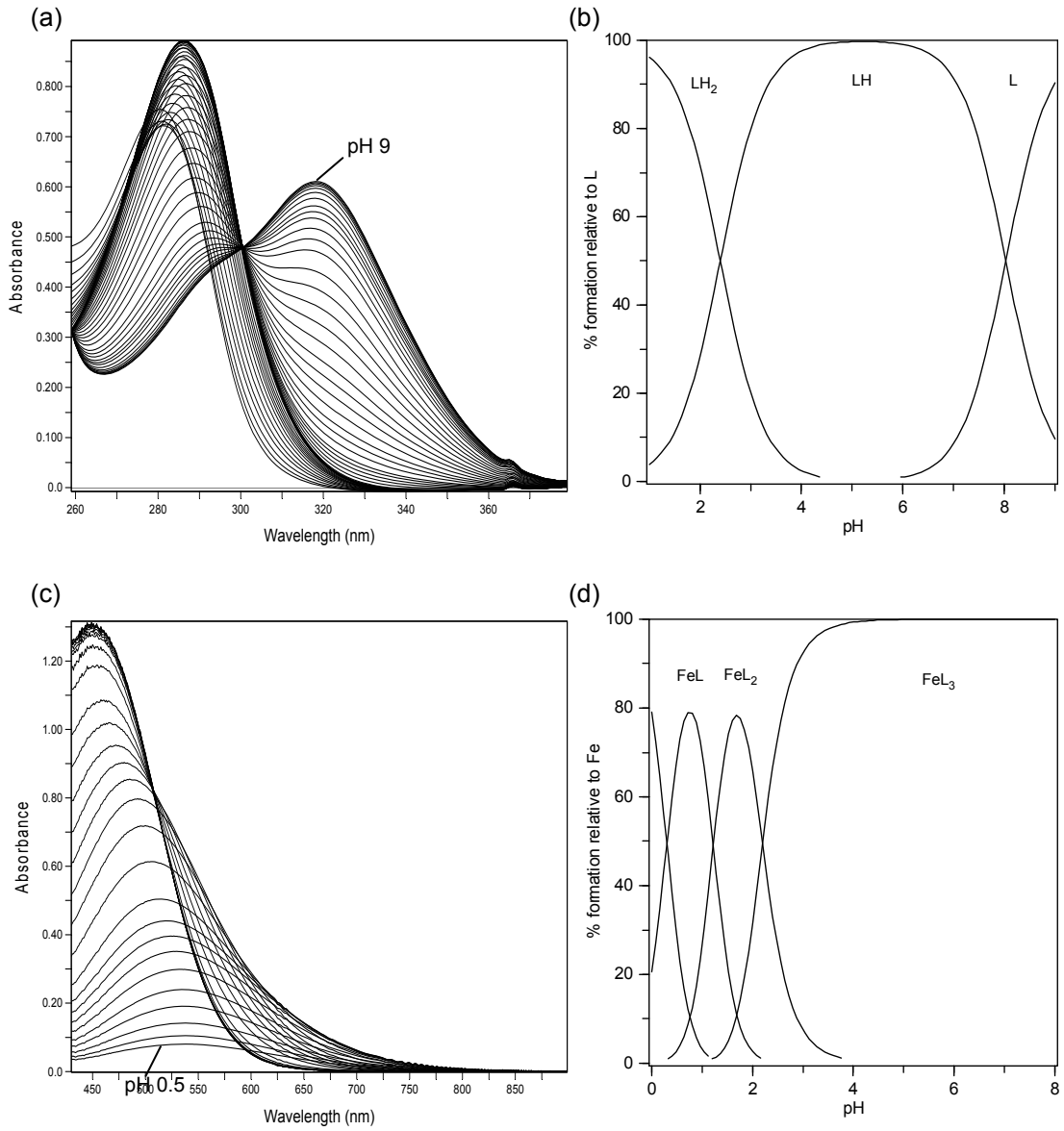


Figure A - 18 UV/Vis spectra and corresponding speciation plots for CP510.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1002.6 µM over the pH range 1-9; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 454.0 µM and [Fe^{3+}] = 44.8 µM over the pH range 0.5-8.

A.19 CP529

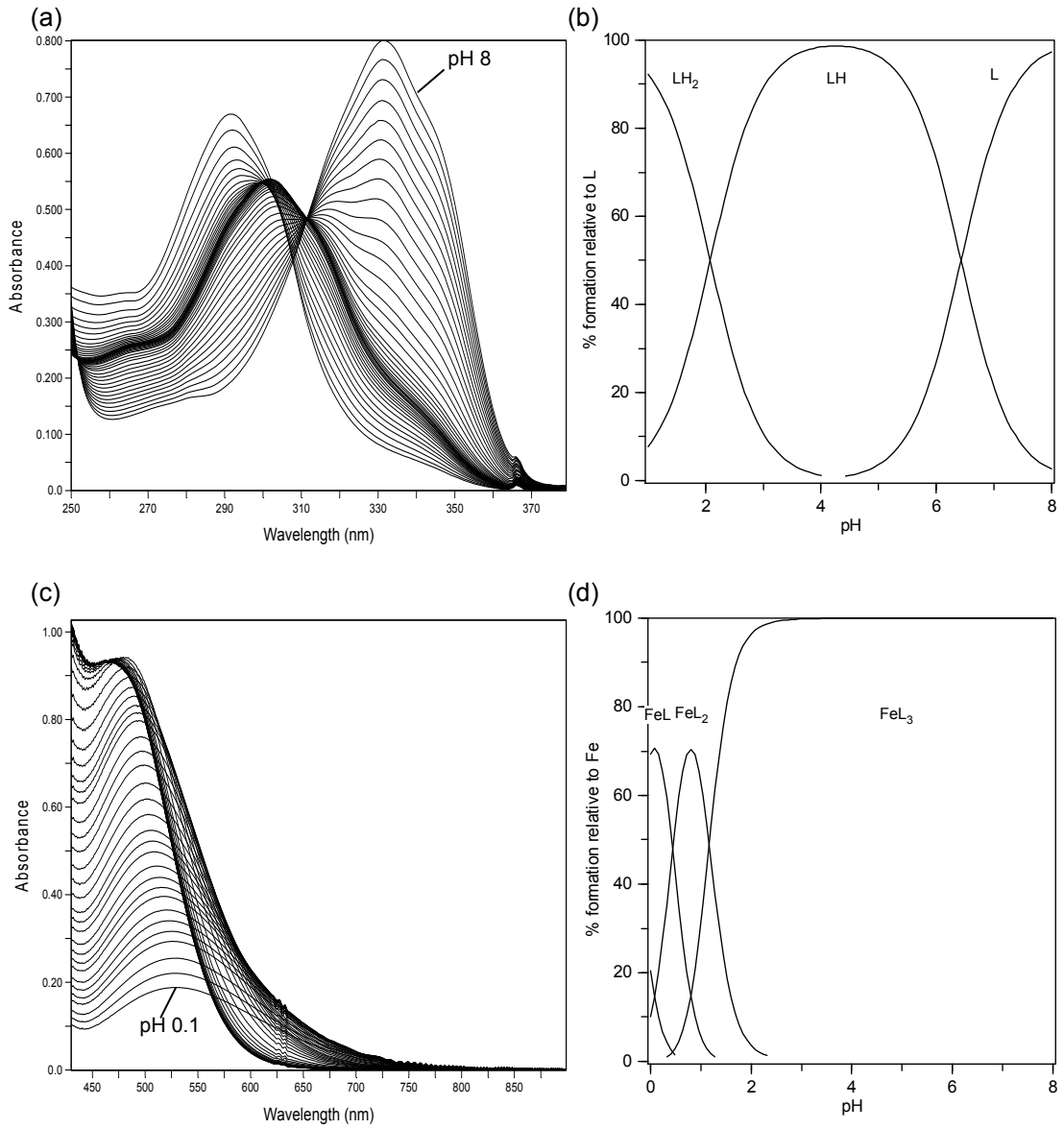


Figure A - 19 UV/Vis spectra and corresponding speciation plots for CP529.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 746.6 µM over the pH range 1-8; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 430.7 µM and [Fe^{3+}] = 42.8 µM over the pH range 0.1-8.

A.20 CP545

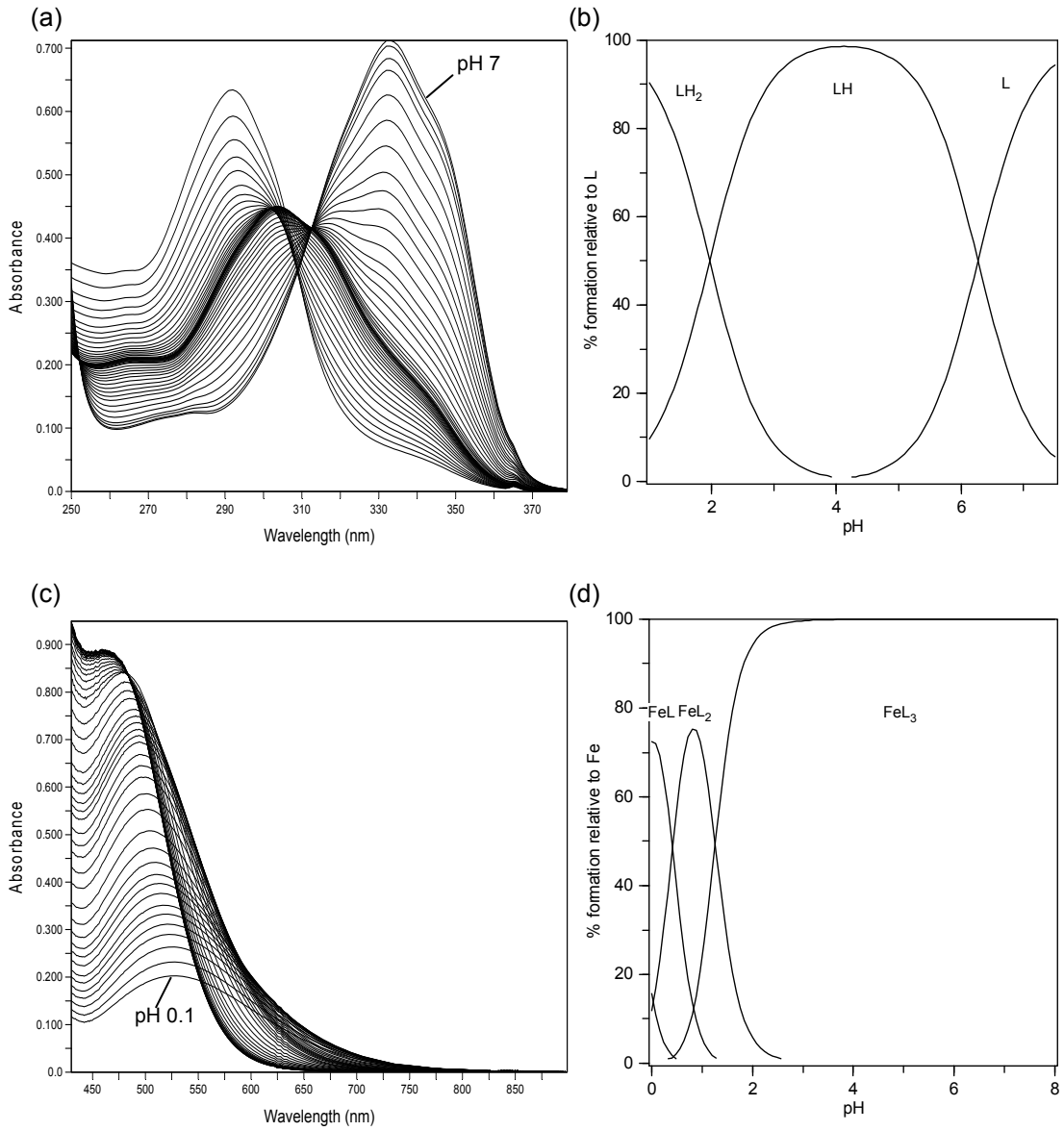


Figure A - 20 UV/Vis spectra and corresponding speciation plots for CP545.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 724.3 µM over the pH range 1-7; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 429.9 µM and [Fe^{3+}] = 42.8 µM over the pH range 0.1-8.

A.21 CP616

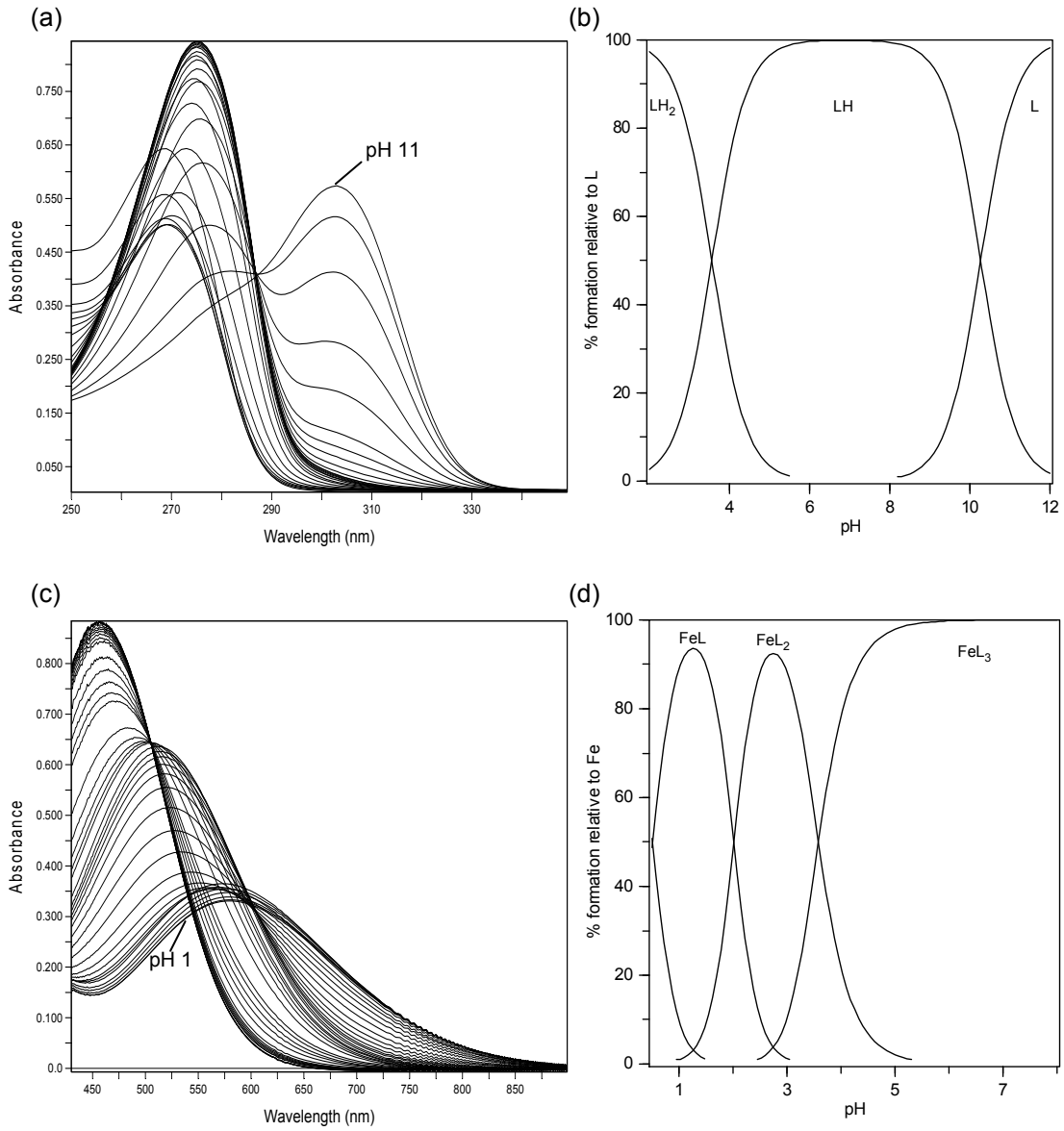


Figure A - 21 UV/Vis spectra and corresponding speciation plots for CP616.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1392.0 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 461.3 µM and [Fe^{3+}] = 46.2 µM over the pH range 1-8.

A.22 CP751

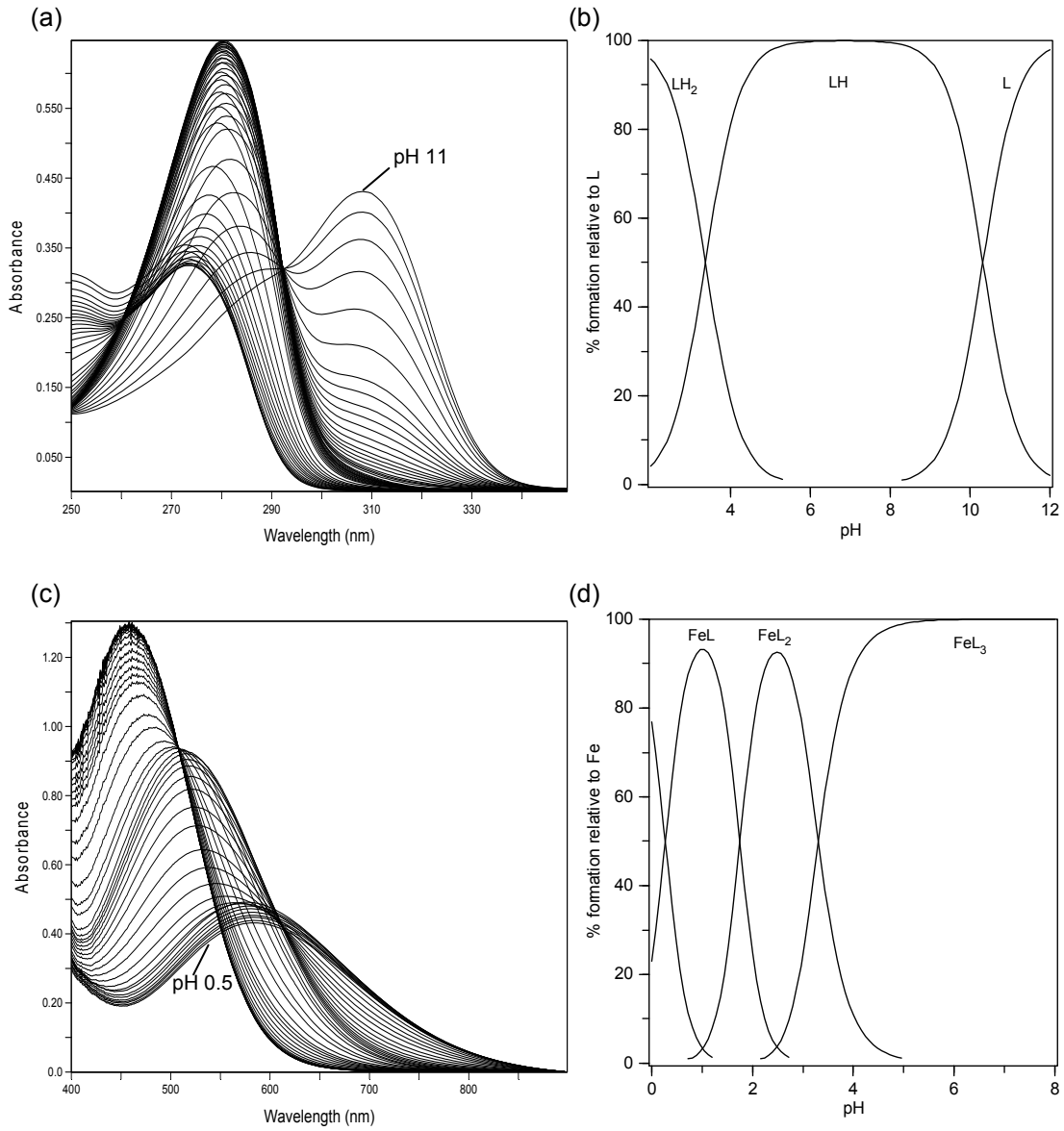


Figure A - 22 UV/Vis spectra and corresponding speciation plots for CP751.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 714.4 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 462.5 µM and [Fe^{3+}] = 46.6 µM over the pH range 0.5-8.

A.23 YMF1

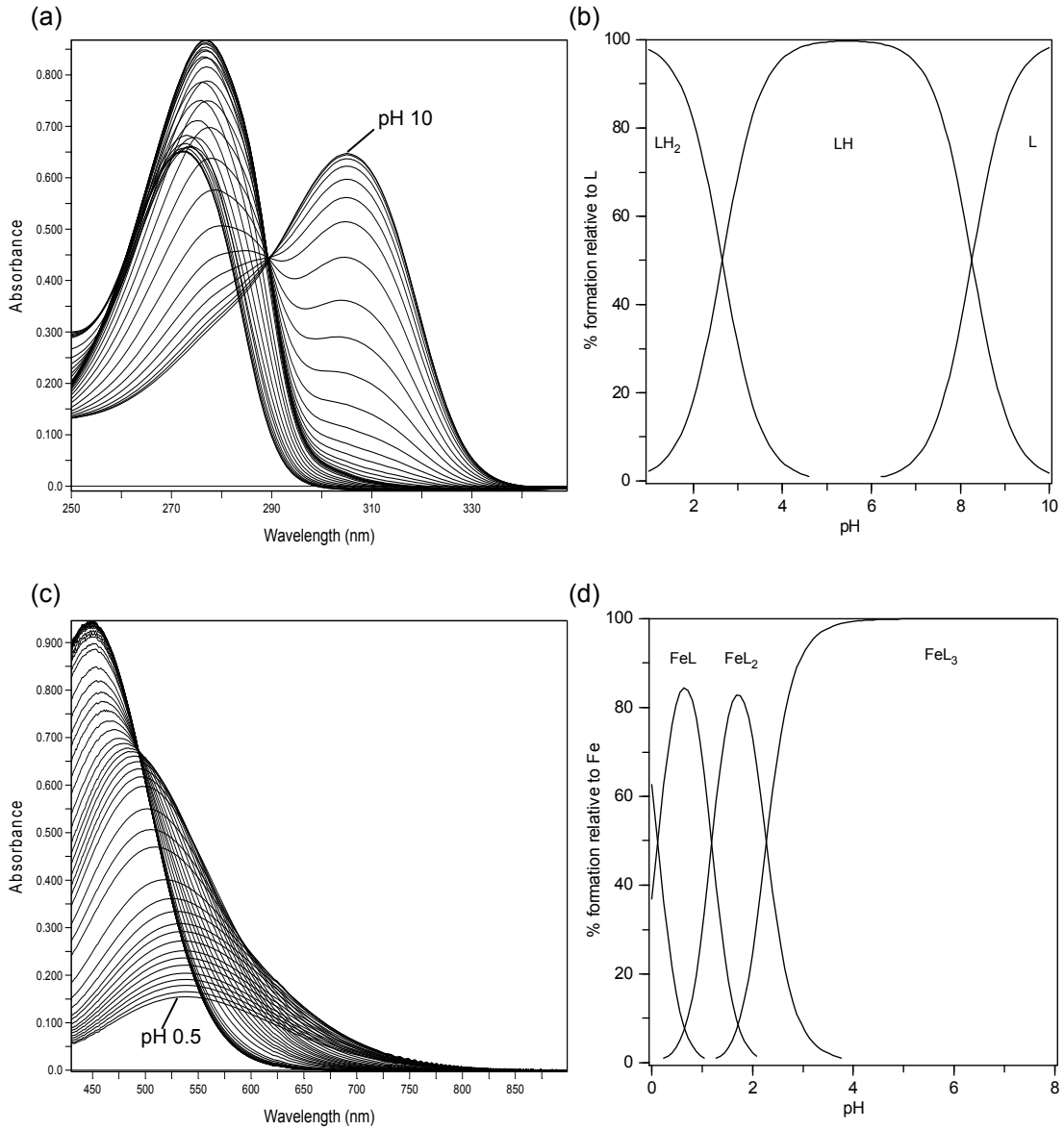


Figure A - 23 UV/Vis spectra and corresponding speciation plots for YMF1.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 831.3 μM over the pH range 2-10; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 444.5 μM and [Fe^{3+}] = 44.8 μM over the pH range 0.5-8.

A.24 YMF3

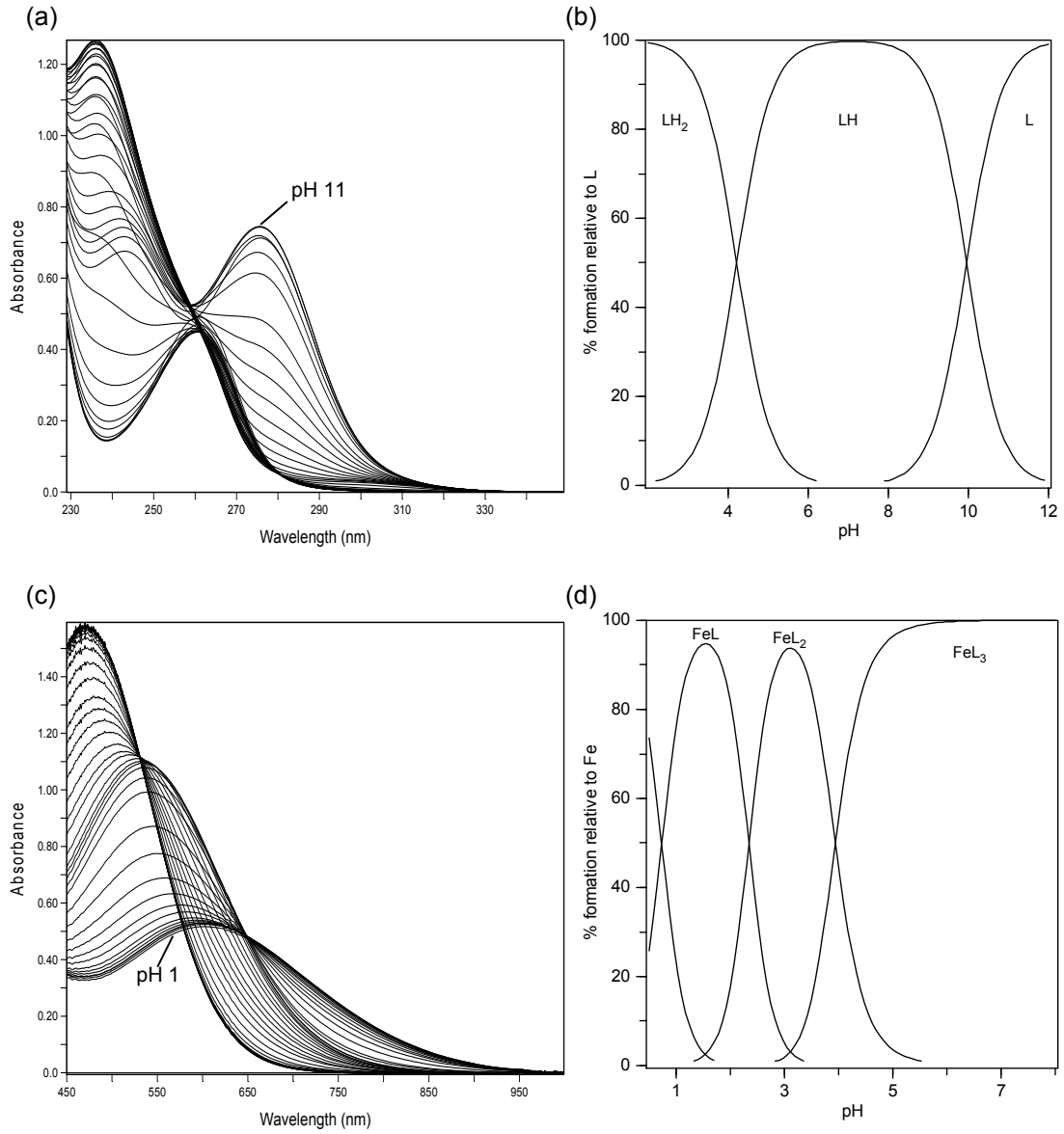


Figure A - 24 UV/Vis spectra and corresponding speciation plots for YMF3.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1278.7 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 460.1 µM and [Fe^{3+}] = 46.8 µM over the pH range 1-8.

A.25 YMF4

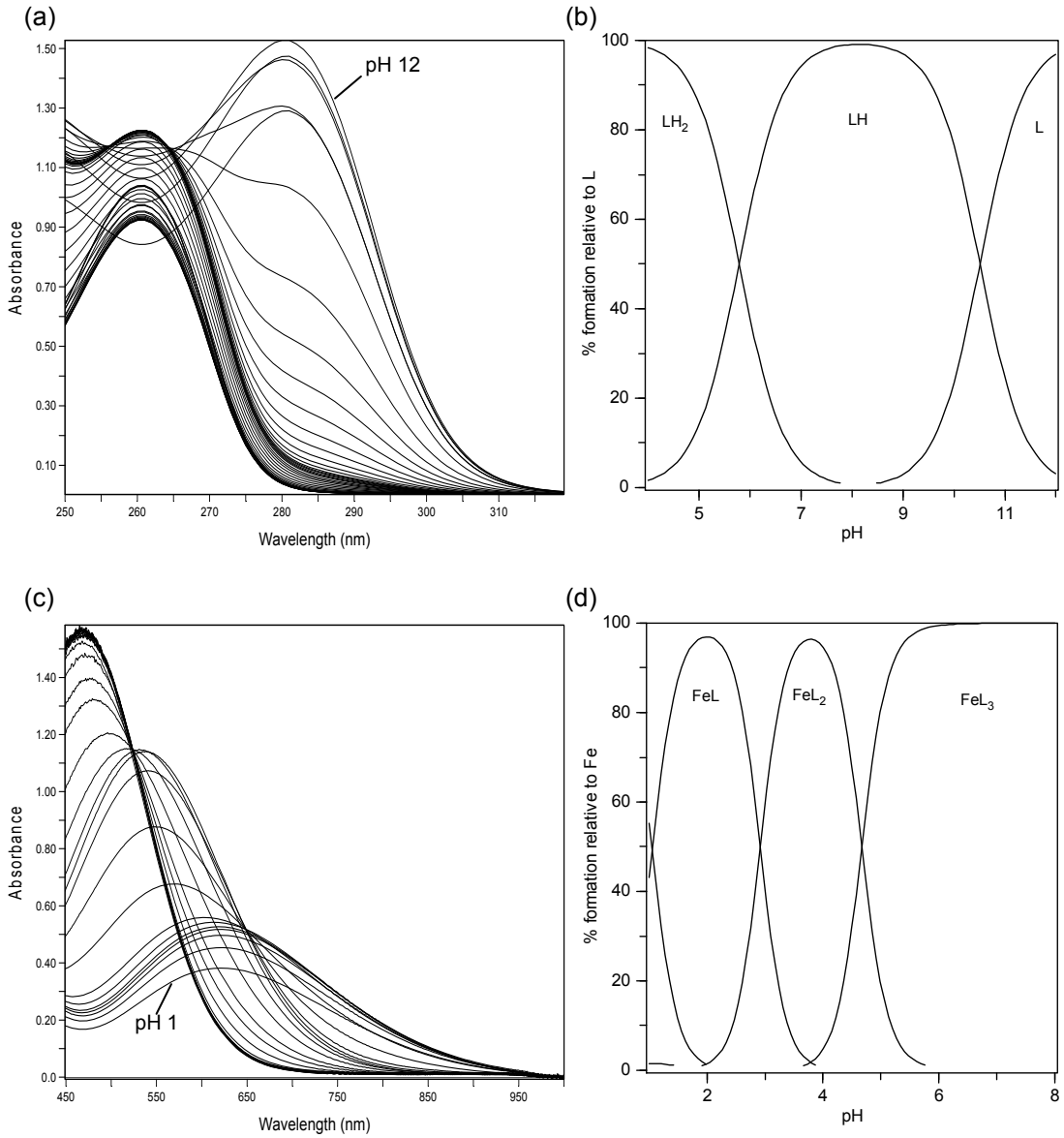


Figure A - 25 UV/Vis spectra and corresponding speciation plots for YMF4.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 2877.5 µM over the pH range 4-12; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 461.1 µM and [Fe^{3+}] = 45.8 µM over the pH range 1-8.

A.26 YMF6

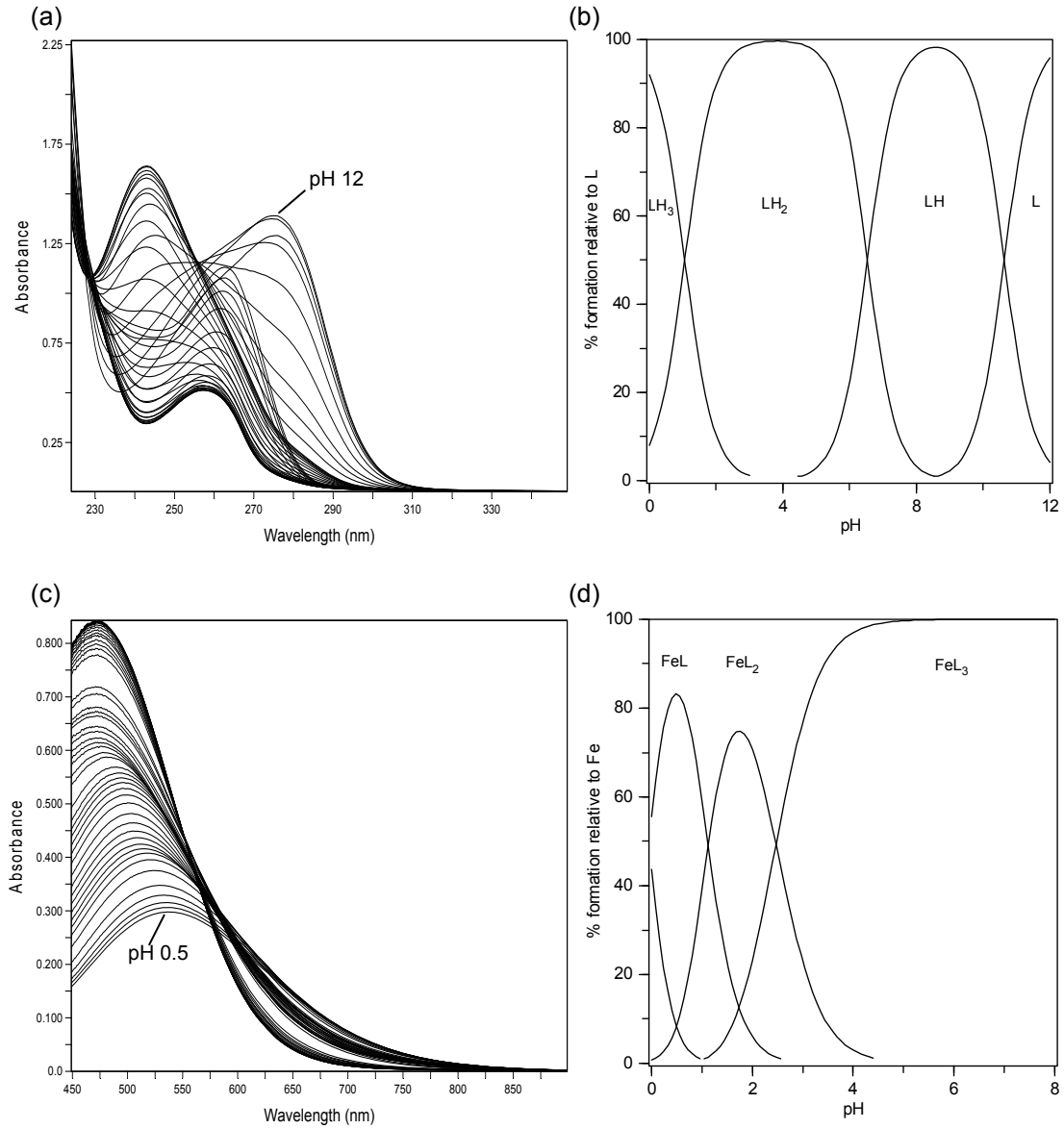


Figure A - 26 UV/Vis spectra and corresponding speciation plots for YMF6.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1538.1 μM over the pH range 0.5-12; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 582.5 μM and [Fe^{3+}] = 56.0 μM over the pH range 0.5-8.

A.27 YMF13

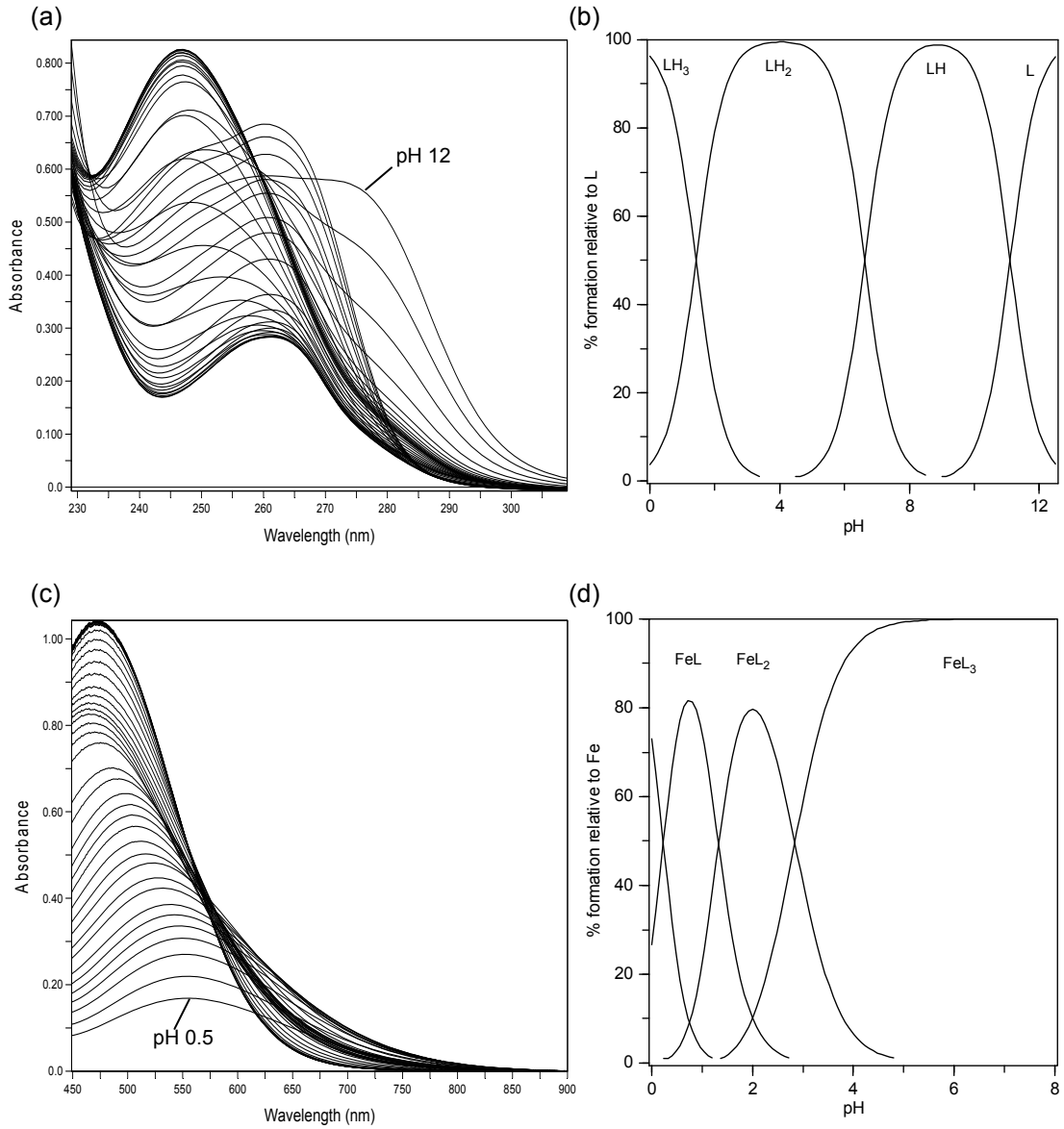


Figure A - 27 UV/Vis spectra and corresponding speciation plots for YMF13.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 1478.6 μM over the pH range 0.5-12; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 458.3 μM and [Fe^{3+}] = 45.3 μM over the pH range 0.5-8.

A.28 YMMO1

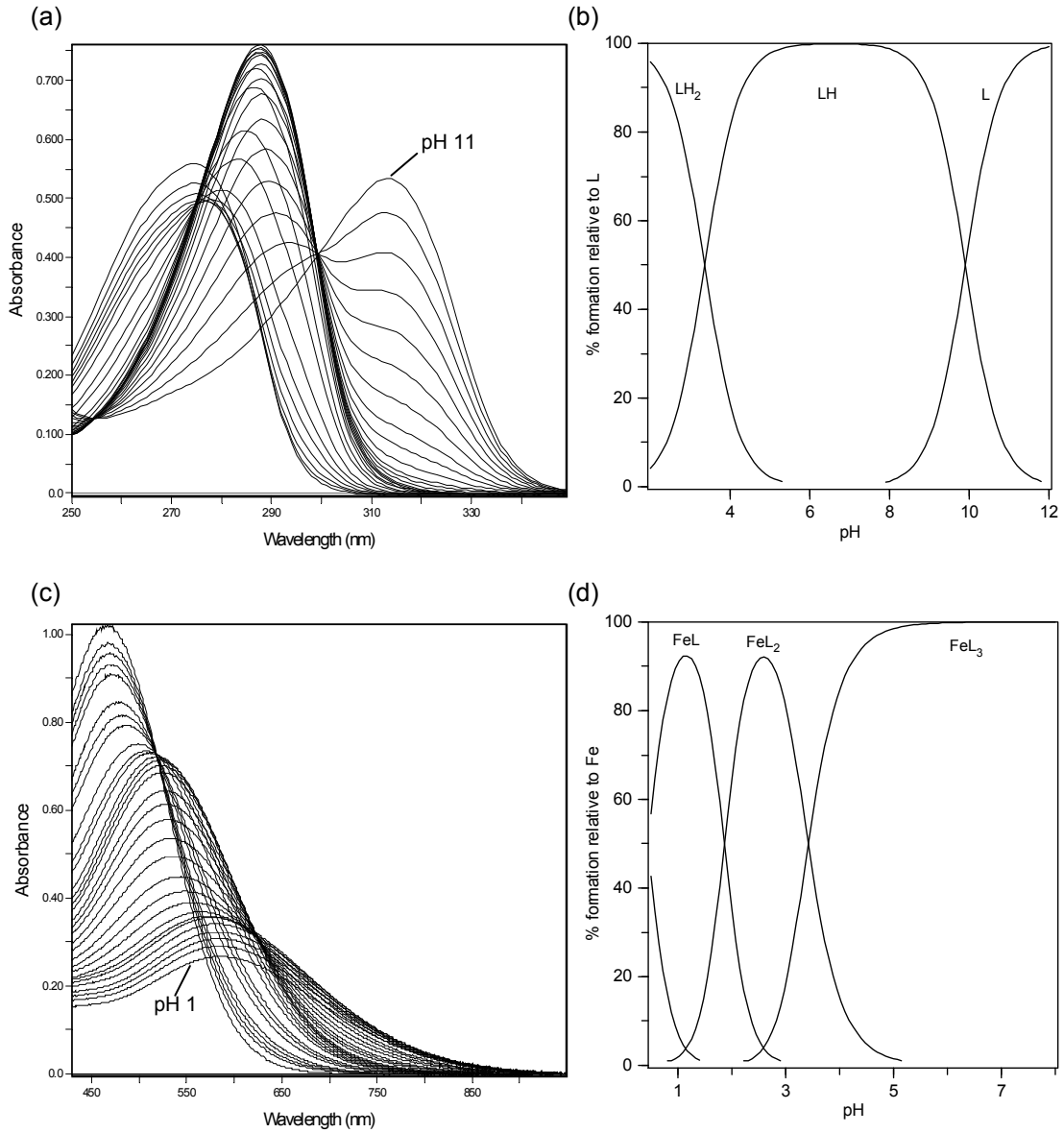


Figure A - 28 UV/Vis spectra and corresponding speciation plots for YMMO1.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 959.4 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 407.2 μM and [Fe^{3+}] = 41.0 μM over the pH range 1-7.

A.29 CN116

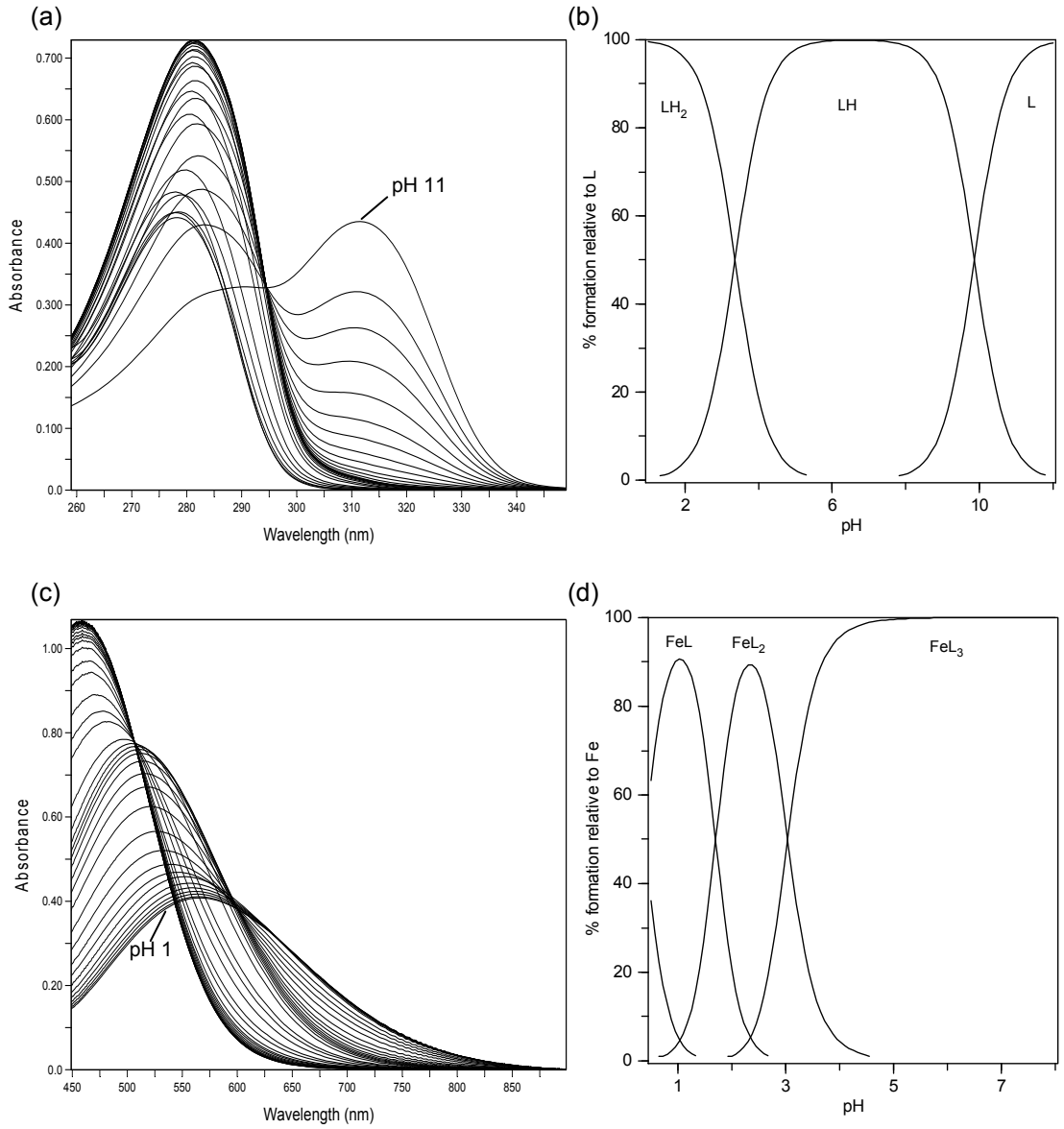


Figure A - 29 UV/Vis spectra and corresponding speciation plots for CN116.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 729.4 µM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 468.2 µM and [Fe^{3+}] = 45.9 µM over the pH range 1-8.

A.30 CN128

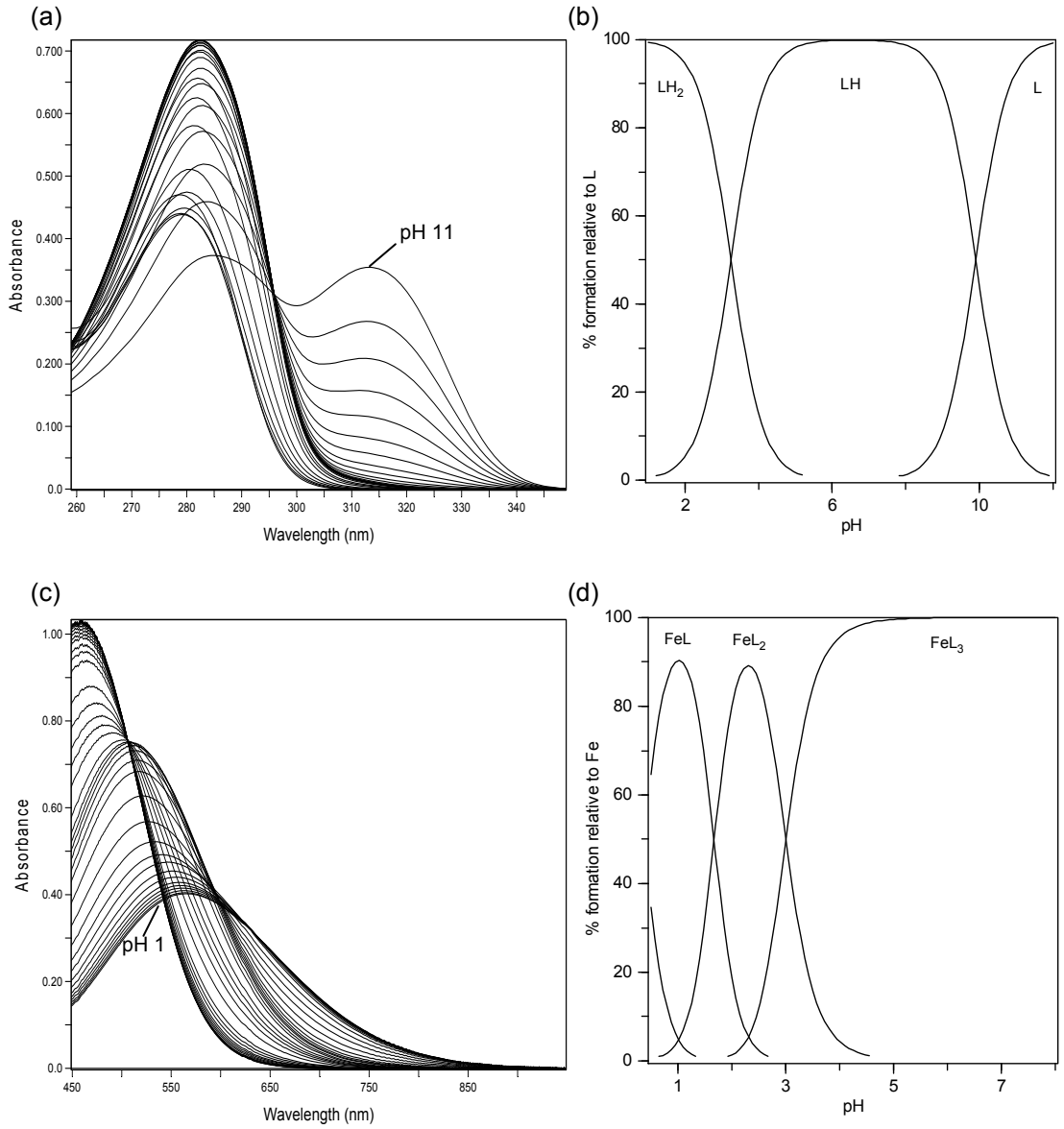


Figure A - 30 UV/Vis spectra and corresponding speciation plots for CN128.

(a)(b): the UV/Vis spectra and corresponding speciation plots for pure ligand when [L] = 728.3 μM over the pH range 2-11; (c)(d): the UV/Vis spectra and corresponding speciation plots for ligand in the presence of Fe^{3+} when [L] = 467.2 μM and [Fe^{3+}] = 45.8 μM over the pH range 1-8.